

PAPER: CLASSICAL STATISTICAL MECHANICS, EQUILIBRIUM AND NON-EQUILIBRIUM

## Origins and diagnostics of the nonequilibrium character of active systems

To cite this article: Lokrshi Prawar Dadhichi *et al* *J. Stat. Mech.* (2018) 123201

View the [article online](#) for updates and enhancements.



**IOP | ebooks™**

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

PAPER: Classical statistical mechanics, equilibrium and non-equilibrium

# Origins and diagnostics of the nonequilibrium character of active systems

Lokrshi Prawar Dadhichi<sup>1</sup>, Ananyo Maitra<sup>2</sup>  
and Sriram Ramaswamy<sup>3</sup>

<sup>1</sup> Tata Institute of Fundamental Research, Gopanpally, Hyderabad 500 107, India

<sup>2</sup> LPTMS, CNRS, University Paris-Sud, Université Paris-Saclay, 91405 Orsay, France

<sup>3</sup> Department of Physics, Centre for Condensed Matter Theory, Indian Institute of Science, Bangalore 560012, India

E-mail: [lpdadhichi@gmail.com](mailto:lpdadhichi@gmail.com), [nyomaitra07@gmail.com](mailto:nyomaitra07@gmail.com) and [sriram@iisc.ac.in](mailto:sriram@iisc.ac.in)

Received 3 September 2018

Accepted for publication 10 October 2018

Published 3 December 2018



Online at [stacks.iop.org/JSTAT/2018/123201](https://stacks.iop.org/JSTAT/2018/123201)  
<https://doi.org/10.1088/1742-5468/aae852>

**Abstract.** We present in detail a Langevin formalism for constructing stochastic dynamical equations for active-matter systems coupled to a thermal bath. We apply the formalism to clarify issues of principle regarding the sources and signatures of nonequilibrium behaviour in a variety of polar and apolar single-particle systems and polar flocks. We show that distance from thermal equilibrium depends on how time-reversal is implemented and hence on the reference equilibrium state. We predict characteristic forms for the frequency-resolved entropy production for an active polar particle in a harmonic potential, which should be testable in experiments.

**Keywords:** active matter, fluctuating hydrodynamics

**Contents**

<b>1. Introduction</b>	<b>3</b>
<b>2. General derivation of active equations</b>	<b>4</b>
<b>3. Applications</b>	<b>6</b>
3.1. Reactive coupling between physical momentum and chemical variable .....	6
3.1.1. Self-propelled dimers.....	6
3.1.2. Field theory of active polar rods on a substrate: Toner–Tu equation.....	7
3.2. Dissipative coupling between physical momentum and chemical variable .....	9
3.2.1. Particle with preferred speed.....	9
3.2.2. Apolar active particle.....	10
3.2.3. Active field theory with velocity alignment.....	11
<b>4. Quantification of nonequilibrium behaviour</b>	<b>13</b>
4.1. Nonequilibrium dynamics of active particles .....	14
4.1.1. HSR when the time-reversal operation does not involve a polarity flip.....	17
4.1.2. HSR when the time-reversal operation involves a polarity flip.....	19
4.1.3. Entropy production of other active particle models.....	21
4.2. Nonequilibrium dynamics of active field theories .....	21
4.2.1. Entropy production for polar rods.....	21
4.2.2. Entropy production for the model with velocity-alignment.....	22
<b>5. Conclusion</b>	<b>22</b>
<b>Acknowledgments</b> .....	<b>23</b>
<b>Appendix A. Active dynamics through the coupling of physical and chemical velocities</b> .....	<b>23</b>
A.1. Reversible coupling of physical and chemical velocities.....	24
A.1.1. Active particle with preferred speed.....	24
A.1.2. Apolar AOUP.....	25
A.1.3. Active field theory with velocity alignment.....	26
<b>Appendix B. Activity through fixed reaction velocity</b> .....	<b>27</b>
<b>Appendix C. Inertial active particles</b> .....	<b>28</b>
<b>Appendix D. Entropy production due to particle driven by pure coloured noise</b> .....	<b>29</b>
<b>Appendix E. Entropy production and time-reversal in general systems with two degrees of freedom</b> .....	<b>30</b>
<b>Appendix F. Nonequilibrium fluctuation-dissipation relation for an active particle in a harmonic trap</b> .....	<b>32</b>

J. Stat. Mech. (2018) 123201

<b>Appendix G. Harada–Sasa relation for active field theory of polar rods .....</b>	<b>34</b>
<b>Appendix H. Entropy production in the homogeneous polarised phase in the <math>T \rightarrow 0</math> limit .....</b>	<b>35</b>
<b>References .....</b>	<b>36</b>

---

## 1. Introduction

Active systems are held away from thermal equilibrium by free energy supplied directly to the constituent particles, which transduce it into systematic movement [1, 2]. This breaking of time-reversal symmetry alone at the scale of the microscopic units, while retaining spatial homogeneity and isotropy, is the defining feature of active matter and sets it apart from familiar driven systems forced by an imposed spatial gradient of temperature, potential or velocity. In this article we construct active-particle dynamics from the Langevin equation for thermal Brownian particles coupled to the chemical kinetics of fuel and offer a fresh perspective on their nonequilibrium character. One aim of this work is to point out that the answer to the question ‘How far from equilibrium is active matter?’ [3] is not unique. It depends on the reference equilibrium state which, in turn, depends on the time-reversal signature assigned to the dynamical variables in question. In addressing this question we show precisely what observed behaviours, as defined by a suitable notion of entropy production, distinguish an active system from a passive counterpart with the same spatial symmetries. A contrast with the treatment of [3] is that in the present work we consider systems in which all degrees of freedom are coupled to a common heat bath at temperature  $T$ .

The first part of our discussion highlights the source of nonequilibrium behaviour. We begin by writing down *equilibrium* Langevin equations for a set of coordinates and momenta coupled to an explicit chemical degree of freedom. We show that the stochastic equations for the dynamics of active particles emerge naturally if the chemical driving force, which is the difference  $\Delta\mu$  between the chemical potentials of reactants and products, is then held fixed. This amounts to an extension of the approach of [4–8] to include noise, inherited in the simplest case from equilibrium Langevin dynamics, but allowing the possibility of strong modification in the presence of driving. In the second part of our treatment we examine the resulting behaviour, that is, how the imposition of a maintained chemical-potential difference leads to entropy production. Note that we define the latter through probability ratios of forward and backward processes with assigned time-reversal signatures [9–11], with no necessary connection to heat generation [12]. This is operationally useful because in experiments one tracks only a few of the many possible degrees of freedom—e.g. the position and vectorial orientation of a single self-propelled particle [13]. We shall see that such partial entropy-production functions do estimate how far the observed dynamics is from thermal equilibrium. Of course, the calculation of entropy production requires the definition of a time-reversal operation, and such a definition should lead to vanishing entropy production rate in the limit of vanishing  $\Delta\mu$ . However, two different equilibrium limits are possible in standard models for active polar

particles—one in which the self-propulsion velocity remains a velocity and another in which it lapses back to being simply the orientation vector. We discuss this both for single polar active particles—active Brownian particles (ABPs) or active Ornstein–Uhlenbeck particles (AOUPs) [3, 14–18]—and the Toner–Tu field theory [19–21] for a collection of such particles. The latter is then seen as the active variant of either a polar liquid crystal or a Navier–Stokes fluid, with different entropy production rates.

Here are our main results.

- All standard stochastic dynamical models for active matter, including nominally two-temperature cases such as those for apolar active particles, as well as those in which activity enters as a negative linear-order damping, emerge naturally from our single-bath framework.
- A non-unit coefficient for the ‘advective nonlinearity’ in the Toner–Tu model arises only if both Galilean invariance and detailed balance are broken.
- Translational Brownian motion in the dynamics of polar active particles leads to a statistical steady state qualitatively different from that obtained in [3, 22], with nonzero entropy production even for a particle in a harmonic potential. The value and form of the entropy production rate depends on whether the time-reversal involution includes an inversion of the polarity. The characteristic profiles we predict for frequency-resolved entropy production rates (see figure 1), from a Harada–Sasa approach [23], should be testable in experiments on granular or colloidal active matter. The predicted profile for the case where the polarisation is treated as even under time-reversal is consistent with that discussed in [24] for Markovian systems with timescale separation.
- An active polar particle with translational diffusion in a harmonic trap is distinguishable from a passive one only when the dynamics of the auxiliary variable is also recorded.

This paper is organised as follows. In section 2 we summarise the approach to the construction of stochastic equations for active systems. In section 3 we apply the formalism to generate the dynamics of a variety of single-particle and spatially extended systems, starting from polar active particles (active Brownian or active Ornstein–Uhlenbeck) as motile dimers. In section 4 we evaluate entropy production rates via time-reversal of the stochastic action, and construct Harada–Sasa relations connecting entropy production to correlation-response differences. We close in section 5 with a summary. A more detailed technical exposition of many relevant points can be found in a series of appendices.

## 2. General derivation of active equations

In this section<sup>4</sup> we review the general derivation of stochastic active dynamics as presented in [26]. In line with [4], we describe active systems as those in which one

<sup>4</sup>The discussion in this and the next sections completely disposes of all issues raised in [25].

or more chemical degrees of freedom drive the observed mechanical degrees of freedom. We construct equations of motion for dynamical variables  $\mathcal{C}$  with position-like and momentum-like components  $\mathcal{Q}$  and  $\mathcal{P}$ , respectively even and odd under time-reversal (hereafter denoted  $\mathcal{T}$ ).  $\mathcal{C}$  can be finite- or infinite-dimensional, depending on whether the system of interest is a single active particle [27], or a spatially extended system described by an active field theory such as [19, 21].  $\mathcal{Q}$  and  $\mathcal{P}$  need not be canonically conjugate variables nor even have the same number of components. The stochastic equations of motion describing the *thermal equilibrium* dynamics of  $\mathcal{C}$  are [28–31]

$$\partial_t \mathcal{C} = -(\mathbf{\Gamma} + \mathbf{W}) \cdot \nabla_{\mathcal{C}} H + T \nabla_{\mathcal{C}} \cdot \mathbf{W} + \xi \quad (1)$$

where  $H(\mathcal{C})$  is the effective Hamiltonian,

$$\langle \xi(t) \xi(t') \rangle = 2T \mathbf{\Gamma} \delta(t - t'), \quad (2)$$

$\mathbf{\Gamma}$  is a symmetric matrix of dissipative couplings between the variables and  $\mathbf{W}$  is an antisymmetric matrix of reactive couplings. Terms involving  $\mathbf{W}$  must have, component by component, the same signature under  $\mathcal{T}$  as  $\partial_t \mathcal{C}$ , and those involving  $\mathbf{\Gamma}$  must have the opposite  $\mathcal{T}$ -signature. Thus the  $\mathcal{Q}\mathcal{Q}$  and  $\mathcal{P}\mathcal{P}$  components of  $\mathbf{\Gamma}$  must themselves be even under  $\mathcal{T}$ , while the  $\mathcal{Q}\mathcal{P}$  and  $\mathcal{P}\mathcal{Q}$  components must be odd. The case of specific interest to us is where  $\mathcal{Q}$  consists of a spatial part  $\mathbf{X}$  and the chemical coordinate  $n$ , and  $\mathcal{P}$  has only a spatial part  $\mathbf{P}$ . Then the components  $\mathbf{\Gamma}_{\mathbf{X}\mathbf{X}}$ ,  $\mathbf{\Gamma}_{\mathbf{X}n}$ ,  $\mathbf{\Gamma}_{\mathbf{P}\mathbf{P}}$ , and  $\Gamma_{nn}$  must be even under  $\mathcal{T}$ , and  $\mathbf{\Gamma}_{\mathbf{X}\mathbf{P}}$  and  $\mathbf{\Gamma}_{\mathbf{P}n}$  must be odd. Since we are not considering the possibility of an external field that breaks  $\mathcal{T}$ , such as a magnetic field [32, 33], this implies that  $\mathbf{\Gamma}_{\mathbf{X}\mathbf{P}}$ , and  $\mathbf{\Gamma}_{\mathbf{P}n}$  should themselves be odd in  $\mathbf{P}$ . When  $\mathbf{\Gamma}$  depends on  $\mathcal{C}$  the noise (2) is multiplicative. For the steady-state distribution to be  $e^{-H/T}$ , we must then include in (1) the additional drift  $T(\nabla_{\mathcal{C}} \cdot \mathbf{\Gamma} - \alpha \mathbf{g} \cdot \nabla_{\mathcal{C}} \mathbf{g})$  where  $\mathbf{g} \cdot \mathbf{g} = 2\mathbf{\Gamma}$  [31], and  $\alpha \in [0, 1]$  parameterizes the noise interpretation. Similarly,  $\mathbf{W}_{\mathbf{P}\mathbf{P}}$  should be odd under  $\mathcal{T}$  and therefore suitably  $\mathbf{P}$ -dependent. The term  $T \nabla_{\mathcal{C}} \cdot \mathbf{W}$  in (1), which emerges in standard derivations of generalized Langevin equations [28, 29, 34, 35], is required for the steady state to be  $e^{-H/T}$ . In familiar Langevin equations [36] for dynamics at equilibrium this derivative vanishes, but in appendices A.1.1 and A.1.3 we will present natural instances where it is nonzero.

In [26], the active terms were obtained by starting from a description in which the physical momentum is entrained by an off-diagonal *dissipative* coupling to a momentum formally conjugate to  $n$ , taking the limit of vanishing inertia for this ‘chemical momentum’, eliminating it in favour of the chemical force, and finally holding the latter fixed at a nonzero value. We demonstrate in appendix A that every active model we consider can be obtained this way as well. Time-reversal symmetry of course dictates that active systems with a *reactive* coupling between  $\mathbf{P}$  and  $n$  can be obtained by the above dissipative entrainment of the physical velocity by the chemical one while those where this coupling is *dissipative* require a reversible, anti-symmetric coupling between the physical and chemical velocities.

Our aim is to arrive at dynamical equations for the mechanical degrees of freedom in the presence of a chemical driving  $\Delta\mu$  in the Hamiltonian which we now write as

$$H = H_0 + n\Delta\mu. \quad (3)$$

To do this, we will now ignore all dissipative terms in the  $X_i$  equations. We will also assume that neither  $\mathbf{\Gamma}$  nor  $\mathbf{W}$  nor  $H_0$  depend on  $n$ . In the absence of driving, the discrete character of  $n$  which counts fuel molecules consumed, and the kinetic barrier for a single chemical reaction, should be incorporated by including in  $H_0$  a term periodic in  $n$ , but we shall ignore this detail in our coarse-grained description. This leads to the simplified active equations of motion

$$\partial_t X_i = -\mathcal{W}_{X_i P_j} \frac{\partial H}{\partial P_j} \tag{4}$$

$$\partial_t P_i = -\Gamma_{P_i P_j} \frac{\partial H}{\partial P_j} - \Gamma_{P_i n} \Delta\mu - \mathcal{W}_{P_i X_j} \frac{\partial H}{\partial X_j} - \mathcal{W}_{P_i n} \Delta\mu - \mathcal{W}_{P_i P_j} \frac{\partial H}{\partial P_j} + \xi_{P_i} \tag{5}$$

where  $\Gamma_{P_i n}$  is odd in  $P_i$  by symmetry and  $\mathcal{W}_{P_i n}$  is even under  $\mathcal{T}$  and therefore, apart from constants, can only involve  $X_i$  and even functions of  $\mathbf{P}$ . The noise correlator now has the usual form

$$\langle \xi_{P_i}(t) \xi_{P_j}(t') \rangle = 2T \Gamma_{P_i P_j} \delta(t - t'). \tag{6}$$

Equations (4) and (5) constitute the stochastic generalisation of the nonequilibrium thermodynamical derivation of the active matter equations proposed in [4]. In [4], however, the effects of interest entered primarily through a reactive coupling of  $n$  to  $\mathbf{P}$ ; we will see that a dissipative coupling between the two quantities also introduces qualitatively distinct physics.

In the next section we use this approach to motivate the dynamical equations both for active particles and active field theories.

### 3. Applications

#### 3.1. Reactive coupling between physical momentum and chemical variable

*3.1.1. Self-propelled dimers.* As our first example, we consider a dimer [27, 37] with centre-of-mass position and momentum  $X$  and  $P$ , and relative coordinate and momentum  $x$  and  $p$ . We do not allow any dissipative coupling between  $P$  and  $p$  and reactive coupling between  $p$  and  $n$ , but assume, in the notation of section 2, a reversible off-diagonal kinetic coefficient  $\mathcal{W}_{Pn} = -\zeta x$ . Then from (4) and (5) we immediately obtain the mechanical equations of motion at a constant chemical driving  $\partial_n H = \Delta\mu$ :

$$\dot{X} = \partial_P H; \dot{x} = \partial_p H \tag{7}$$

$$\dot{P} + \Gamma \partial_P H = -\partial_X H + \zeta \Delta\mu x + \xi_P \tag{8}$$

$$\dot{p} + \gamma \partial_p H = -\partial_x H + \xi_p \tag{9}$$

where we have taken  $\Gamma_{PP} = \Gamma$  and  $\Gamma_{pp} = \gamma$ , and the thermal Gaussian white noises  $\xi_P$  and  $\xi_p$  have strengths  $2k_B T \Gamma$  and  $2k_B T \gamma$  respectively. The chemical coordinate  $n$



of course does not appear in the coupled  $X - x$  dynamics, but makes its presence felt only through the active driving proportional to  $\Delta\mu$ , which cannot be absorbed within any redefinition of the Hamiltonian or the friction coefficient [26]. If, additionally, we assume  $x$  enters  $H$  as  $kx^2/2$ , and drop inertia in both the equations by formally setting masses to zero, we obtain the equation for an active Ornstein–Uhlenbeck particle [3, 38] with  $(\zeta\Delta\mu/\Gamma)x$  as the coloured noise, with correlation time  $\gamma/k$ , in the  $X$  dynamics, but with an additive white noise  $\xi_P$  as well. A centrosymmetric bistable potential for  $x$  with deep wells or, in higher dimensions, a wine-bottle form with a deep trough, is equivalent to the ABP [15]. Within our single-bath stochastic description it is inconsistent to set  $\xi_P$  to 0 without also setting  $\xi_p$  to 0 in (9). The presence of translational diffusion  $\xi_P$  in the dynamics of  $X$  distinguishes our model, and the resulting stationary state, qualitatively from that of [3]. On the other hand, *if we simply abolish  $\xi_P$  in (8), then—still working with masses set to zero—for the case where the internal coordinate  $x$  is harmonically bound,  $\partial_x H = kx$ , this system of equations is equivalent to the second-order stochastic dynamics*

$$\frac{\Gamma\gamma}{k}\ddot{X} = -\left(\Gamma + \frac{\gamma}{k}\partial_X^2 H\right)\dot{X} - \partial_X H + \frac{\zeta\Delta\mu}{k}\xi_P \quad (10)$$

with Gaussian white noise  $\xi_p$ . For the special case in which  $H$  is a quadratic function of  $X$ ,  $\partial_X^2 H = K = \text{constant}$ , and (10) reduces to the dynamics of a Brownian damped harmonic oscillator with mass  $\Gamma\gamma/k$  and spring constant  $K$ , with constant damping coefficient  $\Gamma_{\text{eff}} = \Gamma + \gamma K/k$  at thermal equilibrium at temperature  $T_{\text{eff}} = T(\zeta\Delta\mu/k)^2\gamma/\Gamma_{\text{eff}}$ . In general, as already remarked by [3], (10) represents a nonequilibrium dynamics, whose departure from an effective thermal equilibrium is governed by the anharmonicity of  $H$  and not by  $\Delta\mu$ . However, the above analysis applies only if the additive noise  $\xi_P$  in (8) is set to zero. This limit is inconsistent if the noise terms in both equations of motion originate in a common bath, but is in general permissible and important as a case where the magnitude of the effective temperature is governed by the strength of the activity, in a manner analogous to [39, 40]. Further, this reduces to a different equilibrium model [17, 18] if the  $\ddot{X}$  term on the left-hand side of (10) is ignored, which corresponds to the unified-coloured-noise approximation [41, 42].

*3.1.2. Field theory of active polar rods on a substrate: Toner–Tu equation.* We will now use (4) and (5) as applied to fields to obtain the coarse-grained dynamical equations for a collection of active polar rods on a substrate, thus situating the Toner–Tu theory of flocks [19, 20, 43] in the framework of spatial degrees of freedom forced by chemical driving. The rods are described by fields  $\rho(\mathbf{r}, t)$  for the density,  $\mathbf{g}(\mathbf{r}, t)$  for the momentum density (or velocity  $\mathbf{v} = \mathbf{g}/\rho$ ) and  $\mathbf{w}(\mathbf{r}, t)$  for the polarisation, i.e. the collective vectorial orientation of the rods. The substrate on which the system resides serves as a momentum sink, resulting in a nonzero damping even for a spatially uniform velocity. It also offers a fixed frame of reference, breaking Galilean invariance and thus permitting a constant field-independent contribution to the components  $\mathcal{W}_{w_i(\mathbf{x}), g_j(\mathbf{x}'')}$  of the matrix of reversible kinetic coefficients. The result is a reactive coupling  $\dot{\mathbf{w}} \propto \mathbf{v}$ . To see this heuristically, note that in two-fluid models [44], every phase is advected by its own velocity and not by the velocity of the other phase. A system on a substrate can



be regarded as a two-fluid model, viewed from the frame of reference of the second fluid which is treated as a rigid object.

At a single-particle level, the polarization can be viewed, as in section 3.1, as the relative coordinate of a dimer made of two dissimilar particles. If its centre of mass moves with velocity  $\mathbf{v}$  relative to an ambient medium, which could simply be a fixed substrate, the drag experienced by the two ends will in general be unequal. This differential drag implies that the relative velocity will also experience a drag proportional to the centre-of-mass velocity, whose components along and normal to the dimer stretch/compress and rotate the dimer. The total force on the relative coordinate  $\mathbf{w}$  is then  $-(1/2)(\gamma_1 - \gamma_2)\mathbf{v} - (1/4)(\gamma_1 + \gamma_2)\dot{\mathbf{w}}$  + conservative forces, if the dimer consists of a pair of equal masses with drag coefficients  $\gamma_1$  and  $\gamma_2$ . In a description where inertia is ignored, the rate of change of the polarisation then contains the ‘weathercock effect’ [45, 46], a contribution  $\Lambda\mathbf{v}$  where  $\Lambda = 2(\gamma_1 - \gamma_2)/(\gamma_1 + \gamma_2)$ , although arising in this approach as the ratio of dissipative coefficients, is a reversible kinetic coefficient<sup>5</sup>.

Extending this notion to a many-particle system, the passive equations of motion for polar rods on a substrate, with a free-energy  $H_0 = F[\rho, \mathbf{w}] + \mathbf{g}^2/2\rho$ , are

$$\partial_t \rho = -\nabla \cdot \mathbf{g} \tag{11}$$

$$\partial_t \mathbf{w} + \mathbf{v} \cdot \nabla \mathbf{w} + \boldsymbol{\Omega} \cdot \mathbf{w} = -\Gamma_w \frac{\delta F}{\delta \mathbf{w}} + \Lambda \mathbf{v} + \lambda \mathbf{w} \cdot \mathbf{A} + \boldsymbol{\xi}^w \tag{12}$$

$$\partial_t \mathbf{g} + \nabla \cdot (\mathbf{v}\mathbf{g}) = -\rho \nabla \frac{\delta F}{\delta \rho} - \Gamma \mathbf{v} + (\nabla \mathbf{w}) \cdot \frac{\delta F}{\delta \mathbf{w}} + \nabla \cdot \left[ \mathbf{w} \frac{\delta F}{\delta \mathbf{w}} \right]^A + \lambda \nabla \cdot \left[ \mathbf{w} \frac{\delta F}{\delta \mathbf{w}} \right]^S - \Lambda \frac{\delta F}{\delta \mathbf{w}} + \boldsymbol{\xi}_g \tag{13}$$

where the noise correlations are given by  $\langle \boldsymbol{\xi}^w(\mathbf{x}, t) \boldsymbol{\xi}^w(\mathbf{x}', t') \rangle = 2T\Gamma_w \mathbf{1} \delta(\mathbf{x} - \mathbf{x}') \delta(t - t')$  and

$$\langle \boldsymbol{\xi}_g(\mathbf{x}, t) \boldsymbol{\xi}_g(\mathbf{x}', t') \rangle = 2T\Gamma \delta(\mathbf{x} - \mathbf{x}') \delta(t - t'), \tag{14}$$

and  $\boldsymbol{\Omega}$  and  $\mathbf{A}$  are the anti-symmetric and symmetric parts of the velocity gradient tensor respectively.

As advertised, we now write down the corresponding dynamical equation in the presence of a chemical driving  $\Delta\mu$  in the Hamiltonian, now written as  $H = H_0 + \int_{\mathbf{r}} n(\mathbf{r}) \Delta\mu(\mathbf{r})$ , considering a reactive coupling between the momentum  $\mathbf{g}(\mathbf{r}, t)$  and the chemical variable  $n(\mathbf{r}, t)$ . This appears in the  $\mathbf{g}$  equation through the term

$$\int_{\mathbf{r}'} \mathcal{W}_{\mathbf{g}(\mathbf{r}), n(\mathbf{r}')} \frac{\delta H}{\delta n(\mathbf{r}')}, \tag{15}$$

where the vector  $\mathcal{W}_{\mathbf{g}(\mathbf{r}), n(\mathbf{r}')}$  has to be even under time-reversal. Therefore, to the lowest order in gradients,

$$\mathcal{W}_{\mathbf{g}(\mathbf{r}), n(\mathbf{r}')} = -\zeta_w \mathbf{w}(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}') \tag{16}$$

<sup>5</sup> In a bulk polar liquid crystal [47] momentum conservation rules out this coupling, allowing only  $\nabla^2 \mathbf{v}$  at leading order in gradients and fields. The reader may wonder how the Poisson bracket—in which the reversible coupling originates—of the same pair of variables change by introducing a substrate. The answer is that it’s not the same pair of variables. The velocity with respect to the substrate is like the relative velocity of the two components of a binary fluid.

where  $\zeta_w$  is a phenomenological kinetic coefficient<sup>6</sup>.

Inserting the coupling (15) and (16) into (13), creating an active state by requiring  $\delta H/\delta n(\mathbf{r}) = \Delta\mu$ , i.e. imposing a spatially uniform, constant, nonzero reactant-product chemical-potential difference  $\Delta\mu$ , and discarding inertia, we get

$$\Gamma \mathbf{v} = \zeta_w \Delta\mu \mathbf{w} - \nabla \rho - \Lambda \frac{\delta F}{\delta \mathbf{w}} + \boldsymbol{\xi}_g, \quad (17)$$

so that the continuity equation (11) becomes

$$\partial_t \rho = -\nabla \cdot \left( \frac{\zeta_w \Delta\mu}{\Gamma} \rho \mathbf{w} \right) + \frac{\Lambda}{\Gamma} \nabla \cdot \left( \rho \frac{\delta F}{\delta \mathbf{w}} \right) - \nabla \cdot \left[ \frac{\rho}{\Gamma} (-\nabla \rho + \boldsymbol{\xi}_g) \right] \quad (18)$$

and the polarisation equation is

$$\begin{aligned} \partial_t \mathbf{w} = & \left( \frac{(\lambda - 1)\zeta_w}{2\Gamma} \right) \Delta\mu \mathbf{w} \cdot \nabla \mathbf{w} - \frac{\lambda}{\Gamma} \mathbf{w} \cdot \nabla \nabla \rho + \left( \frac{(\lambda - 1)\zeta_w^2}{4\Gamma} \right) \Delta\mu \nabla \mathbf{w}^2 \\ & + \frac{1}{\Gamma} \nabla \rho \cdot \nabla \mathbf{w} - \frac{\Lambda}{\Gamma} \nabla \rho + \left( \frac{\Lambda \zeta_w}{\Gamma} \right) \Delta\mu \mathbf{w} - \left( \Gamma_w + \frac{\Lambda^2}{\Gamma} \right) \frac{\delta F}{\delta \mathbf{w}} + \boldsymbol{\xi} \end{aligned} \quad (19)$$

where the noise correlator, to the zeroth order in gradients, is  $\langle \boldsymbol{\xi}(\mathbf{x}, t) \boldsymbol{\xi}(\mathbf{x}', t') \rangle = 2T [\Gamma_w + (\Lambda^2/\Gamma)] \delta(\mathbf{x} - \mathbf{x}') \delta(t - t')$ . We see thus that our general framework generates coupled equations for the polar order parameter  $\mathbf{w}$  and the density  $\rho$  with the two characteristic features of the Toner–Tu equation [19, 20]: 1. A self-advective nonlinearity and 2. A number current proportional to the polarisation *with an independent coefficient*. We discuss in greater detail in section 3.2.3 why the simultaneous presence of a self-advective coefficient and a particle current proportional to polarisation with an independent coefficient inevitably breaks detailed balance. Note that the noise in the polarisation and concentration equation are not independent since a part of the noise in the  $\mathbf{w}$  equation is derived from the noise in the momentum equation. However, this cross-correlation is  $\mathcal{O}(\nabla)$  and turns out to be irrelevant compared to the wavevector-independent noise in the  $\mathbf{w}$ —the correct low frequency and small wavevector scaling of all correlation functions can be obtained by completely ignoring the noise in the density equation (except perpendicular to the wavevector, but in that case  $\mathbf{w}$  and  $\rho$  are linearly decoupled).

### 3.2. Dissipative coupling between physical momentum and chemical variable

*3.2.1. Particle with preferred speed.* Distinct from the AOUP we discussed in section 3.1, a particle can be active by possessing a preferred speed obtained through the interplay of a negative linear damping and a positive nonlinear one [15, 48, 49]. We now show how a dissipative coupling between the physical momentum and the chemical coordinate can create a realisation of such a particle. Working for simplicity in one

<sup>6</sup> At first order in gradients and with one field, we only have  $\nabla \rho$ . At first order in gradients and two fields, we have three terms with two  $\mathbf{w}$  and one gradient, and three terms with two  $\mathbf{v}$  and one gradient. Terms like  $\mathbf{w} \cdot \nabla \mathbf{v}$  are also allowed by the spatial symmetries of this system which lacks Galilean invariance, but by time-reversal symmetry *can not* arise from a reactive coupling between  $n$  and  $\mathbf{g}$ . Instead, they arise through the dissipative coupling between  $n$  and  $\mathbf{g}$ ,  $-\Gamma_{gn} \delta H/\delta n$ .  $\Gamma_{gn}$  can contain all vectors containing an odd power of  $\mathbf{v}$  and, in particular, contains terms such as  $\mathbf{w} \cdot \nabla \mathbf{v}$ .

dimension, consider a particle with position and momentum  $X, P$ , with equation of motion

$$\dot{X} = P; \quad \dot{P} = -\partial_X H - \Gamma_{PP}P - \Gamma_{Pn}\partial_n H + \xi_P = -\partial_X H - \Gamma_{PP}P + \zeta\Delta\mu P + \xi_P \quad (20)$$

in the notation of section 2, where we have taken  $\Gamma_{Pn} = -\zeta P$ . If we now take  $\Gamma_{PP} = \gamma_1 + \gamma_2 P^2$ , we obtain

$$\dot{P} = -\partial_X H - (\gamma_1 - \zeta\Delta\mu + \gamma_2 P^2)P + \xi_P. \quad (21)$$

Note that the  $P$ -independent part of the friction is now liberated from the noise strength because of the  $\Delta\mu$ -dependent term, which can even make it negative. The noise is multiplicative with the correlator  $\langle \xi_P(0)\xi_P(t) \rangle = 2T(\gamma_1 + \gamma_2 P^2)\delta(t)$ . This implies that (21) must be interpreted using the anti-Itô interpretation. In a general interpretation, one must add a term  $2T\gamma_2(1 - \alpha)P$  to the R.H.S. of (21), where  $\alpha$  parameterizes the interpretation, with  $\alpha = 0$  for Itô,  $1/2$  for Stratonovich, and  $1$  for anti-Itô [31]. Despite superficial appearances, however, this noise-induced drift can not turn the  $P$ -independent part of the friction negative. In all interpretations, the probability distribution of  $P$  is Maxwellian if  $\Delta\mu = 0$ , and the particle is not self-propelled. If  $\zeta > \gamma_1/\Delta\mu$ , the zero-velocity state is on average unstable, and the particle moves with preferred velocity  $\pm\sqrt{(\zeta\Delta\mu - \gamma_1)/\gamma_2}$  in the limit of low noise. We have thus succeeded in producing a model self-propelled particle with inertia within our consistent stochastic framework.

The reader may wonder whether the requirement of equilibrium dynamics when a nonzero average  $\partial_n H$  is not imposed places any restriction on the magnitude of  $\zeta$ . It does not: the coupled dissipative dynamics of  $P$  and  $n$  is given by

$$\partial_t \begin{pmatrix} P \\ n \end{pmatrix} = \begin{pmatrix} \gamma_1 + \gamma_2 P^2 & -\zeta P \\ -\zeta P & \Gamma_{nn} \end{pmatrix} \begin{pmatrix} \partial_P H \\ \partial_n H \end{pmatrix} + \begin{pmatrix} \xi_P \\ \xi_n \end{pmatrix} + \text{NID}$$

where NID denotes noise-induced drift. The friction matrix must be positive definite. This yields the condition  $\Gamma_{nn}(\gamma_1 + \gamma_2 P^2)/P^2 > \zeta^2$ . This condition is obviously fulfilled for small  $P$ , while at large  $P$  it requires  $\Gamma_{nn}\gamma_2 > \zeta^2$ . However,  $\Gamma_{nn}$  does not enter into the equation for  $P$  and therefore, this condition does not pose any restriction for  $\zeta$ —it simply provides a lower limit to the friction coefficient in the  $n$  equation which becomes immaterial once  $\partial_n H$  is externally held constant. The noise-induced drift that is required for this system to be in equilibrium can also be calculated from this (see earlier footnote). After a little algebra, we see that the noise-induced drift in the  $P$  equation is, as expected,  $2T\gamma_2(1 - \alpha)P$ .

*3.2.2. Apolar active particle.* Macroscopic self-driven systems with alignment but no polarity—active nematics—have received a great deal of attention in the literature [50], but studies of the single-particle behaviour have focused on polar particles. A single apolar active particle executes uncorrelated random back-and-forth movement preferentially along an axis, as seen in certain elongated cells [51] or in a vertically vibrated rod lying on a surface [46]. This amounts to the kinetic energy being unequally partitioned between the components of the velocity along and transverse to the rod.

We show that this apparently two-temperature dynamics emerges naturally from our single-bath framework.

Consider a particle with position  $\mathbf{X}$ , momentum  $\mathbf{P}$  and intrinsic anisotropy characterised by the apolar tensor  $\mathbf{Q}$ , and chemical coordinate  $n$ , governed by a Hamiltonian  $H$ . We take a simple relaxational model for  $\mathbf{Q}$ :

$$\dot{\mathbf{Q}} = -\Gamma_{QQ} \frac{\delta H}{\delta \mathbf{Q}} + \xi_Q \quad (22)$$

with  $\langle \xi_Q(0) \xi_Q(t) \rangle = 2T\Gamma_{QQ} \delta(t)$ , reducing to simple rotational diffusion in the limit where  $\delta H / \delta \mathbf{Q} \propto \mathbf{Q}$ .  $\mathbf{Q}$  can enter the dissipative cross-kinetic coefficient between  $\mathbf{P}$  and  $n$  in combination with  $\mathbf{P}$ :

$$\dot{\mathbf{X}} = \mathbf{P}; \quad \dot{\mathbf{P}} = -\partial_{\mathbf{X}} H - \Gamma_{PP} \partial_{\mathbf{P}} H - \Gamma_{Pn} \partial_n H + \xi_P = -\partial_{\mathbf{X}} H - \Gamma_{PP} \partial_{\mathbf{P}} H - \zeta \Delta \mu \mathbf{Q} \cdot \mathbf{P} + \xi_P \quad (23)$$

where we have taken  $\Gamma_{Pn} = \zeta \mathbf{Q} \cdot \mathbf{P}$ . This implies that the momentum is damped anisotropically while the noise correlator  $\langle \xi_P(t) \xi_P(t') \rangle = 2T\Gamma_{PP} \delta(t - t')$  is isotropic. In other words, this is like having an anisotropic temperature at the scale of individual particles [52], yet achieved within our isothermal framework where the driving is imposed only through  $\Delta \mu$ . The inertia-free limit of the dynamics, more familiar from active nematic theories [53], can be obtained by adiabatic elimination of  $\mathbf{P}$  from (22) and (23). We will explore the phase behaviour and other statistical properties of collections of such particles elsewhere<sup>7</sup>.

*3.2.3. Active field theory with velocity alignment.* In section 3.1.2, we had derived Toner–Tu equations for a system of polar rods, whose orientation vector dictated their preferred velocity via the active motility which arose as a reactive kinetic coefficient between the momentum density and the chemical variable. However, there are driven systems without an independent polarity where nevertheless a spontaneous transition occurs to a state with non-zero mean velocity. A particularly striking example of this is the zero-resistance state in microwave-driven 2DEG systems in a magnetic field [54, 55]. The Toner–Tu-like theory of such a system [56] is best viewed as an active extension of the Navier–Stokes equations (modified by contact with a substrate), not an active version of polar liquid crystal dynamics, as there is no underlying orientational degree of freedom.

As in (21), the crucial ingredient in such a theory is a negative linear friction which has to enter as a off-diagonal dissipative coupling between the momentum density field  $\mathbf{g}(\mathbf{r}, t)$  and  $n(\mathbf{r}, t)$ . That is, the standard Navier–Stokes equation (together with substrate damping) is modified as

$$\partial_t \rho = -\nabla \cdot \mathbf{g} \quad (24)$$

$$\partial_t \mathbf{g} + \nabla \cdot (\mathbf{v} \mathbf{g}) = -\rho \nabla \frac{\delta F}{\delta \rho} - \Gamma_{gg} \mathbf{v} + [\eta \nabla^2 + (\zeta + \eta/3) \nabla \nabla \cdot] \mathbf{v} - \Gamma_{gn} \frac{\delta H}{\delta n} + \xi_g \quad (25)$$

<sup>7</sup> Note that though the kinetic coefficient explicitly depends on  $\mathbf{P}$ , which implies that the noise cross-correlation in the coupled  $n - \mathbf{P}$  equation is multiplicative, there is no noise induced drift in either equation even when there is no external driving force on  $n$ . This is due to the symmetric traceless character of  $\mathbf{Q}$ .

with  $H = F[\rho] + \int_{\mathbf{x}} \mathbf{g}^2/2\rho + n\Delta\mu$ , where  $F$  is the purely  $\rho$ -dependent part of the free-energy, and  $\langle \boldsymbol{\xi}_g(\mathbf{r}, t)\boldsymbol{\xi}_g(\mathbf{r}', t') \rangle = 2T\mathbf{I}[\Gamma_{gg} - \eta\nabla^2 - (\zeta + \eta/3)\nabla\nabla]\delta(\mathbf{r} - \mathbf{r}')\delta(t - t')$ . Again, following (21), we take  $\Gamma_{gg} = \gamma_1 + \gamma_2\mathbf{v} \cdot \mathbf{v}$  and take<sup>8</sup>  $\Gamma_{gn} = -\zeta\mathbf{v}$ .

The nonlinear damping  $\propto \gamma_2$  again implies that the noise is multiplicative and (25), as it stands, has to be interpreted using an anti-Itô interpretation. In any other interpretation, a noise-induced drift  $2(1 - \alpha)T\mathbf{v}/\rho$  has to be added. Note however, that unlike the  $\zeta$  term, which when the chemical field is driven with a constant force, i.e.  $\partial_n H = \Delta\mu$ , is liberated from noise strength, this term is intimately connected to the noise and is required precisely to ensure that the steady state distribution of velocities remains Maxwellian in the equilibrium limit irrespective of the noise interpretation. Thus, while  $\zeta > \gamma_1/\Delta\mu$  linearly destabilises the zero velocity state, the noise-induced drift has no such effect. The fluid acquires a preferred mean speed with respect to the substrate in this case which in the low temperature limit is  $[(\zeta\Delta\mu - \gamma_1)/\gamma_2]^{1/2}$ . This therefore is an active variant of the dynamics of a fluid in contact with a momentum sink or substrate [59]. It may be argued that this is still distinct from the Toner–Tu equation since the coefficient of the advective terms in the equations for number and momentum density still has a coefficient 1. The modification of this requires a *further active* contribution, which due to time-reversal symmetry of the coupled unforced  $\mathbf{g} - n$  dynamics, has to enter not through the dissipative coupling at all but via  $\mathcal{W}_{\mathbf{g}(\mathbf{r})n(\mathbf{r}' )}$  (see (15)), which by symmetry, can contain all even vectors and, in particular, can contain terms such as  $\lambda[\nabla \cdot (\mathbf{g}\mathbf{v})](\mathbf{r})\delta(\mathbf{r} - \mathbf{r}')$  which leads to the term  $\lambda\Delta\mu\nabla \cdot (\mathbf{v}\mathbf{g})$  in the  $\mathbf{g}$  equation.

One may wonder whether a non-unit advective coefficient has to necessarily break detailed balance. After all in equilibrium, the advection term in the momentum equation arises from the reactive  $\mathcal{W}_{g_i(\mathbf{r})g_j(\mathbf{r}' )}$  coupling. The coefficient of this coupling is forced to be 1 because of Galilean invariance (and momentum conservation). However, for fluids on substrates, which neither conserve momentum nor have Galilean invariance, it may be argued that the antisymmetric coupling  $\mathcal{W}_{g_i(\mathbf{r})g_j(\mathbf{r}' )}$ , where  $\mathbf{g}$  is now the momentum *relative to the substrate*, can have a non-unit coefficient  $\lambda$ . This would still result in a stochastic differential equation with a current-free steady-state with a distribution  $e^{-H/T}$  i.e. this would not break detailed balance. However, looking at the form of  $\mathcal{W}_{g_i(\mathbf{r})g_j(\mathbf{r}' )}$  in detail [60], we find that beyond the advective term, it also generates another term  $\propto \rho\nabla(\mathbf{v}^2/2)$ . In standard hydrodynamics, this second term (whose presence in the equations of motion would have implied a lack of momentum conservation and Galilean invariance) is *exactly* cancelled by a term arising from the  $\mathcal{W}_{\mathbf{g}(\mathbf{r})\rho(\mathbf{r}' )}$  coupling. Therefore, multiplying the reactive coupling  $\mathcal{W}_{g_i(\mathbf{r})g_j(\mathbf{r}' )}$  by  $\lambda$ , without modifying the  $\mathcal{W}_{\mathbf{g}(\mathbf{r})\rho(\mathbf{r}' )}$  coupling leads to an extra term  $(\lambda - 1)\rho\nabla(\mathbf{v}^2/2)$  in addition to an advective nonlinearity  $\lambda\nabla \cdot (\mathbf{v}\mathbf{g})$ . The *absence* of this extra nonlinearity with a related coefficient (with the relation being preserved under renormalisation) leads to the breaking of detailed balance. If we seek to remove this second nonlinearity with a symmetry-related coefficient by also multiplying the  $\mathcal{W}_{\mathbf{g}(\mathbf{r})\rho(\mathbf{r}' )}$  coupling by  $\lambda$ , we obtain again obtain a model which obeys detailed balance but in which the  $\nabla \cdot (\mathbf{v}\mathbf{g})$  and the

<sup>8</sup> At next order in gradients,  $\Gamma_{gn}$  admits a contribution proportional to  $\nabla^2\mathbf{v}$  which, if it enters with a negative prefactor, implies a negative effective viscosity coefficient, spontaneously generating velocity gradients [57], controlled at larger wavenumbers by ‘hyperviscosity’, i.e. a stabilising damping of order  $\nabla^4$ . The result is a vector variant of the Swift–Hohenberg equation [58] which, when the linear friction coefficient is also negative, presumably displays regular modulated states such as stripes, as well as spatiotemporal chaos.

$\rho \nabla \delta F / \delta \rho$  terms in the momentum equation and the  $\nabla \cdot \mathbf{g}$  term in the density are all multiplied by the *same* coefficient  $\lambda$ <sup>9</sup>. Thus, velocity and density advection at different speeds, in the absence of other symmetry-related nonlinearities, requires breaking of detailed balance.

It must be emphasised however that a model with a negative linear friction and *unit* advective coefficient<sup>10</sup> still belongs to the same universality class as Toner–Tu [20, 21, 43] what is crucial is not that the advective coefficient can have an arbitrary value, but that it is present at all in a model that can spontaneously break symmetry. For such a system, any non-zero value of the advective nonlinearity, including 1, leads to long range order in two dimensions. Moreover, in at least the ‘Malthusian’ [61] variant of Toner–Tu in two dimensions, this vertex is actually protected from renormalisation by an exact symmetry (pseudo-Galilean invariance) which is analogous to the Galilean symmetry that protects this vertex in the Navier–Stokes equation.

#### 4. Quantification of nonequilibrium behaviour

In this section we will discuss how to quantify the degree of nonequilibriumness in some of the models discussed in the last section. To accomplish this, we define the entropy production rate as

$$\sigma = \lim_{t \rightarrow \infty} \frac{1}{t} \mathcal{S} \quad (26)$$

where the entropy production [62]

$$\mathcal{S} = \left\langle \ln \frac{\mathcal{P}}{\mathcal{P}^R} \right\rangle \quad (27)$$

is given by the Kullback–Leibler (KL) divergence [63] that measures the distinguishability of the probability weight associated with a noise realisation to generate a forward trajectory  $\mathcal{P}$  and the weight  $\mathcal{P}^R$  of the noise-realisation required to obtain the time-reversed counterpart of that trajectory. The angle brackets denote an average over noise realisations. However, under suitable ergodicity assumptions, which we implicitly use throughout the paper, the average over noise realisations can be replaced by the average over a single infinitely long noise realisation and, therefore, the angle brackets can be dropped [3, 64].

The entropy production rate that we have defined distinguishes a forward trajectory from its time-reversal counterpart. In equilibrium, time-reversal symmetry implies that the two should be indistinguishable. Therefore,  $\sigma$  provides a measure of how far this system is out of equilibrium. This measure crucially depends on the definition of the time-reversal operation, however. In most cases the requirement that the dynamics

<sup>9</sup> If in addition, the friction coefficient  $\Gamma_{gg}$  is set to 0, this model possesses Galilean invariance and conserves momentum. That is, multiplying both the  $\mathcal{W}_{g_i(\mathbf{r})g_j(\mathbf{r}'')}$  and the  $\mathcal{W}_{\mathbf{g}(\mathbf{r})\rho(\mathbf{r}'')}$  couplings by  $\lambda$  does not spoil either momentum conservation or Galilean invariance.

<sup>10</sup> A model in which the mass current proportional to  $\mathbf{g}$ , the advective nonlinearity and the pressure term are all multiplied by  $\lambda$  or one in which the advective nonlinearity  $\lambda \nabla \cdot (\mathbf{v}\mathbf{g})$  is accompanied by  $(\lambda - 1)\rho \nabla \cdot (\mathbf{v}^2/2)$  also belong to the same universality class.



in the limit  $\Delta\mu = 0$  should have detailed balance fixes the definition of this operation but, when distinct time reversal schemes give zero entropy production in the  $\Delta\mu = 0$  limit, more than one equilibrium limit exists (appendix E). Therefore, in such cases, the answer to the question ‘how far from equilibrium is an active system’ is not unique

#### 4.1. Nonequilibrium dynamics of active particles

We will now quantify the degree of nonequilibriumness of active polar particles discussed in section 3.1. The first step in calculating the entropy production rate and all the subsequent analysis consists in writing the path probability of the particle. Since the noises in both (8) and (9) are Gaussian, the path probability weight [65] takes the form  $\mathcal{P} = e^{-\mathcal{A}}$ , where, for the inertia-less limit of the equations of motion (8) and (9), the stochastic action in Onsager–Machlup [65, 66] form is

$$\mathcal{A} = \int_{t_i}^{t_f} dt \left[ \frac{1}{4T\Gamma} \left( \Gamma\dot{X} + \frac{\partial H}{\partial X} - vx \right)^2 + \frac{1}{4T\gamma} (\gamma\dot{x} + kx)^2 - \alpha \left( \frac{\partial^2 H}{\partial X^2} + 1 \right) \right], \quad (28)$$

where we have defined  $v \equiv \zeta\Delta\mu$  and taken  $\partial H/\partial x = kx$ . The term with  $\alpha \in [0, 1]$  comes from the Jacobian for the variable transformation from the noise  $(\xi_P, \xi_p)$  to  $(X, x)$ , with  $\alpha = 0, 1/2, 1$  corresponding respectively to the Itô, Stratonovich and anti-Itô discretisations. Let us define  $\mathcal{T}_1$  to be the time-reversal operation in the normal sense of that term, i.e. without a flip of the polarity or relative coordinate, i.e.  $\mathcal{T}_1 x = x; \mathcal{T}_1 X = X$

The weight associated with the time reverse of this, when the time-reversal operation  $\mathcal{T}_1$  does not involve a polarity-flip, i.e.  $\mathcal{T}_1 x = x; \mathcal{T}_1 X = X$ , is given by the action

$$\mathcal{T}_1 \mathcal{A} = \mathcal{A}_R = \int_{t_i}^{t_f} dt \left[ \frac{1}{4T\Gamma} \left( -\Gamma\dot{X} + \frac{\partial H}{\partial X} - vx \right)^2 + \frac{1}{4T\gamma} (-\gamma\dot{x} + kx)^2 - (1 - \alpha) \left( \frac{\partial^2 H}{\partial X^2} + 1 \right) \right]. \quad (29)$$

Therefore, the entropy production is

$$\mathcal{S}_1 = \mathcal{A}_R - \mathcal{A} = -\frac{1}{T} [H(t_i) - H(t_f)] + \frac{v}{T} \int_{t_i}^{t_f} dt \dot{X} x \quad (30)$$

where to obtain the last equality we have used the generalised Itô formula for the time derivative, which for any function  $G[q_i(t)]$ , where  $q_i(t)$  obey first-order dynamics with an additive white noise, is  $\dot{G} = \dot{q}_i \partial_{q_i} G + \frac{(1-2\alpha)}{2} D_{ij}(q_k) \partial_{q_i} \partial_{q_j} G$ , where  $D_{ij}$  is the noise correlation matrix. Since the interpretation does not change under  $\mathcal{T}_1$  for  $\alpha = 1/2$ , and the Itô formula reduces to the standard chain rule, we will only use the Stratonovich interpretation from now on. However, our results are, of course, invariant under the change of interpretations. We now use (26) to calculate the entropy production rate. Since  $H$  is bounded, the contribution from it vanishes in  $\sigma$ . Replacing the time-average in the second term in (30) with an average over the stationary measure we obtain

$$\sigma_1 = \frac{v}{T} \langle \dot{X} x \rangle. \quad (31)$$

For a potential quadratic in  $X$ ,  $\partial H/\partial X = KX$ , this reduces to



$$\sigma_1 = \frac{v^2}{\Gamma k + \gamma K}. \quad (32)$$

If one defines a time-reversal operation  $\mathcal{T}_2$  that does involve a polarity flip, i.e.  $\mathcal{T}_2 X = X; \mathcal{T}_2 x = -x$ , the time-reversed action is

$$\mathcal{T}_2 \mathcal{A} = \mathcal{A}_R = \int_{t_i}^{t_f} dt \left[ \frac{1}{4T\Gamma} \left( -\Gamma \dot{X} + \frac{\partial H}{\partial X} + vx \right)^2 + \frac{1}{4T\gamma} (-\gamma \dot{x} + kx)^2 - (1 - \alpha) \left( \frac{\partial^2 H}{\partial X^2} + 1 \right) \right] \quad (33)$$

from which we can calculate the entropy production rate

$$\sigma_2 = \frac{v}{\Gamma T} \left\langle \frac{\partial H}{\partial X} x \right\rangle. \quad (34)$$

For  $H$  quadratic in  $X$ , this yields

$$\sigma_2 = \frac{v^2 K \gamma}{k \Gamma (\Gamma k + \gamma K)}. \quad (35)$$

As pointed out in [9], the sum of the two entropy productions is a potential independent constant  $v^2/(k\Gamma)$ . Note that (32) and (35) coincide when the relaxation time of  $X$  equals the correlation time of  $x$ , i.e. when  $\Gamma/K = \gamma/k$ . Intriguingly, this is precisely the condition of critical damping for the equivalent damped harmonic oscillator limit of (10). Note that both definitions of the time-reversal operation,  $\mathcal{T}_1$  and  $\mathcal{T}_2$  lead to a non-vanishing entropy production for a quadratic  $H$ , in contrast to [3]. The reason, as we have already pointed out, is that the centre of mass  $X$  is in general subjected to an additive white noise, so that the particle samples a statistical steady state rather different from that in the special case considered in [3]. However, it is important to note that this non-zero entropy production only emerges when one considers the *joint* path probability distribution of  $X$  and  $x$ . We now show that the entropy-production rate vanishes for the dynamics in a quadratic potential when only the dynamics of  $X$  is recorded. To see this let us rewrite the dynamics for  $X$  as  $\Gamma \dot{X} = -\partial_X H + \theta$  where  $\theta$  is a zero-mean Gaussian noise with the correlation  $\langle \theta(t)\theta(s) \rangle = 2T\Upsilon(t-s)$ . The path measure is determined by the kernel  $\mathcal{G}$  defined by  $\int ds \mathcal{G}(t-s)\Upsilon(s-r) = \delta(t-r)$ . We now define a stochastic action over an infinite time-interval  $-\infty < t < \infty$ :

$$\mathcal{A} = \frac{1}{4T} \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} ds \theta(s) \mathcal{G}(t-s) \theta(t) = \frac{1}{4T} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \theta_{-\omega} \mathcal{G}_\omega \theta_\omega. \quad (36)$$

For a process driven by a combination of two uncorrelated noise sources, one white and the other coloured, we take  $\theta = \eta + vx$  where  $\eta$  and  $x$  are uncorrelated to each other and obey  $\langle \eta_\omega \eta_{\omega'} \rangle = 2T\Gamma 2\pi \delta(\omega + \omega')$  and

$$\langle x_\omega x_{\omega'} \rangle = \frac{2T\gamma}{k^2 + \gamma^2 \omega^2} 2\pi \delta(\omega + \omega'). \quad (37)$$

However, now we only record the dynamics of  $X$  and not the auxiliary variable. The noise correlation is

$$2T\Upsilon_\omega = 2T \left( \Gamma + \frac{v^2\gamma}{k^2 + \gamma^2\omega^2} \right) \tag{38}$$

and

$$\mathcal{G}_\omega = \left( \frac{k^2 + \gamma^2\omega^2}{\Gamma k^2 + \Gamma\gamma^2\omega^2 + v^2\gamma} \right). \tag{39}$$

We now subtract  $1/\Gamma$  ( $= \mathcal{G}_\omega|_{v=0}$ ) from  $\mathcal{G}_\omega$  to obtain

$$\tilde{\mathcal{G}}_\omega \equiv \mathcal{G}_\omega - \frac{1}{\Gamma} = -\frac{v^2\gamma/\Gamma^2}{\gamma^2\omega^2 + k^2 + v^2\gamma/\Gamma}. \tag{40}$$

We will henceforth only consider  $\tilde{\mathcal{G}}_\omega$  since it must contain the entire entropy production. We now expand it in a series in  $\omega^2$ :

$$\tilde{\mathcal{G}}_\omega = \frac{v^2\gamma}{\Gamma^2} \left[ -\left( \frac{\Gamma}{\Gamma k^2 + v^2\gamma} \right) + \gamma^2\omega^2 \left( \frac{\Gamma}{\Gamma k^2 + v^2\gamma} \right)^2 - \gamma^4\omega^4 \left( \frac{\Gamma}{\Gamma k^2 + v^2\gamma} \right)^3 + \dots \right]. \tag{41}$$

This expansion can not, in general, be naturally truncated at any order, and is therefore only of formal value. If we assume that derivatives of the fields at all orders vanish as  $\rightarrow \pm\infty$ , we can transform to the field coordinates (we will ignore the Jacobian contribution since that remains even in time) and to real time to obtain the action

$$\mathcal{A} = \frac{1}{4T} \frac{v^2\gamma}{\Gamma^2} \int dt \left[ -\left( \frac{\Gamma}{\Gamma k^2 + v^2\gamma} \right) (\Gamma\dot{X} + \partial_X H)^2 + \gamma^2 \left( \frac{\Gamma}{\Gamma k^2 + v^2\gamma} \right)^2 \left\{ \partial_t (\Gamma\dot{X} + \partial_X H) \right\}^2 - \gamma^4 \left( \frac{\Gamma}{\Gamma k^2 + v^2\gamma} \right)^3 \left\{ \partial_t^2 (\Gamma\dot{X} + \partial_X H) \right\}^2 + \dots \right]. \tag{42}$$

Noting that under the time-reversal operation  $\mathcal{T}$ ,  $\mathcal{T}X \rightarrow X$ , and all total derivative terms vanish since the fields and all their derivatives vanish at the boundaries, we find (after repeated integration by parts) the KL divergence

$$\mathcal{T}\mathcal{A} - \mathcal{A} = \frac{v^2\gamma}{T\Gamma} \int dt \left[ \gamma^2 \left( \frac{\Gamma}{\Gamma k^2 + v^2\gamma} \right)^2 (\partial_t^3 X \partial_X H) + \gamma^4 \left( \frac{\Gamma}{\Gamma k^2 + v^2\gamma} \right)^3 (\partial_t^5 X \partial_X H) + \dots \right]. \tag{43}$$

Now, if  $H$  is harmonic in  $X$ ,  $\partial_X H \propto X$ . We see that in this case the KL divergence vanishes term by term. Each term is of the form  $(\partial_t^{2n+1} X)X$ , where  $n = 1, 2, 3, \dots$ . After integrating by parts  $n$  times, this becomes  $\partial_t^{n+1} X \partial_t^n X = (1/2)\partial_t(\partial_t^n X)^2$ . This vanishes since  $\partial_t^n X$  vanishes at  $t \rightarrow \pm\infty$ . Therefore, the entropy production due to a particle in a harmonic trap vanishes when the dynamics of the auxiliary variable is not kept track of. This is the case even if we explicitly consider translational diffusion. More generally, this result holds for any arbitrary (Gaussian) noise.  $\mathcal{G}_\omega$  can always be expanded in the form of (41) as  $\mathcal{G}_\omega = \nu_1 + \nu_2\gamma^2\omega^2 + \nu_3\gamma^4\omega^4 + \dots$  and the later steps follow trivially, with only the coefficient of each term now being  $\nu_i$ .

This validates our assertion that if one wants to know whether a particle in a harmonic trap is in equilibrium or not, one must record the dynamics of the auxiliary variable. Recording just the position of the particle is insufficient to distinguish

between equilibrium and nonequilibrium dynamics<sup>11</sup>. This also remains true even if one considers underdamped particles or particles driven by arbitrarily coloured Gaussian noise (appendix C).

While we have calculated the expressions for the entropy production rate of a single polar particle, it is conceptually interesting to relate it to the violation of a fluctuation-dissipation relation. Further, since correlation and response functions are quantities that are accessible and relevant to experiment, such a relation would provide a way to directly measure the entropy production rate. Thus, we now calculate a Harada–Sasa relation (HSR) for the entropy production [3, 23, 64, 67–69]. This additionally allows us to access a *frequency-resolved* entropy production which directly shows us the timescales associated with the processes contributing to entropy production.

*4.1.1. HSR when the time-reversal operation does not involve a polarity flip.* We first derive the HSR for the case in which the time-reversal operation does not involve a polarity flip. We use the MSRJD [70–75] formalism, which is particularly well adapted to calculating response functions, to write the action in (28) in the Stratonovich interpretation<sup>12</sup>.

$$\mathcal{A} = \int dt \left[ T(\Gamma\tilde{X}^2 + \gamma\tilde{x}^2) - \frac{1}{2} \left( \frac{\partial^2 H}{\partial X^2} + 1 \right) - i\tilde{X} \left( \Gamma\dot{X} + \frac{\partial H}{\partial X} - vx \right) - i\tilde{x} (\gamma\dot{x} + kx) \right] \quad (44)$$

where we have introduced the response fields  $\tilde{X}$  and  $\tilde{x}$ .

Letting  $H \rightarrow H - h(t)X(t)$ , and proceeding as in [31], we see that the response function

$$\mathcal{R}_{XX}(t - t') \equiv \left. \frac{\delta \langle X(t) \rangle}{\delta h(t')} \right|_{h=0} = \frac{1}{2T\Gamma} \left\langle X(t) \left[ \Gamma\partial_{t'}X(t') + \frac{\partial H_0}{\partial X}(t') - vx(t') \right] \right\rangle_0 \quad (45)$$

where the subscript 0 on the the angle brackets implies that the average is computed at  $h = 0$ . Let us now perform the exchange  $t \leftrightarrow t'$ , and subtract the resulting expression from the original. Note: the difference between  $\langle X(t)(\partial H/\partial X)(t') \rangle_0$  and its  $t \leftrightarrow t'$  counterpart vanishes, regardless of the form of  $H(X)$ , for  $t \rightarrow t'$ , the limit of interest below. If  $H$  is quadratic in  $X$  it vanishes for all  $t$  and  $t'$ . Anticipating the  $t \rightarrow t'$  limit that we will take below, we therefore neglect this term and write

$$\frac{1}{2T\Gamma} [\Gamma \langle X(t)\partial_{t'}X(t') - X(t')\partial_tX(t) \rangle_0 + v \langle X(t')x(t) - X(t)x(t') \rangle_0] = \mathcal{R}(t, t') - \mathcal{R}(t', t). \quad (46)$$

The  $\langle X(t)X(t') \rangle$  correlation function is invariant under time translations. Using this, differentiating with respect to  $t$  and passing to the limit  $t \rightarrow 0$ , we get

$$-\frac{\Gamma}{T} \lim_{t \rightarrow 0} \partial_t [T(\mathcal{R}(t) - \mathcal{R}(-t)) + \partial_t C_{XX}(t)] = \frac{v \langle \dot{X}x \rangle}{T} = \sigma_1 \quad (47)$$

<sup>11</sup> Another equivalent way to obtain this result is to marginalise the distribution

$\mathcal{P}(X, x) = e^{-\mathcal{A}} = \mathcal{P}_X \mathcal{P}_x \mathcal{P}_{Xx} = e^{-(\mathcal{A}_X + \mathcal{A}_x + \mathcal{A}_{Xx})}$  with respect to  $x$  i.e. obtain  $\mathcal{P}(X)|_x = \mathcal{P}_X \langle \mathcal{P}_{Xx} \rangle_{\mathcal{P}_x}$  and use this to calculate the entropy production.

<sup>12</sup> Note however, that the MSRJD and OM actions are completely equivalent and one can go from the OM action to the MSRJD one via a Hubbard–Stratonovich transformation while one can obtain the OM action from the MSRJD one by simply integrating out the response fields.

where  $C_{XX}(t)$  is the  $XX$  correlator, and the last equality follows from (31). Note that if we had retained the  $\langle X(t)(\partial H/\partial X)(t') \rangle_0$  term, it would have vanished at this stage for any  $H$ . Writing the Fourier transform of  $(\mathcal{R}(t) - \mathcal{R}(-t))$  as  $2i\chi''_{XX}(\omega)$ , we get, in the frequency domain,

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{\omega\Gamma}{T} [\omega C_{XX}(\omega) - 2T\chi''_{XX}(\omega)] = \sigma_1 = \int_{-\infty}^{\infty} \sigma_1^\omega d\omega. \quad (48)$$

Explicitly, for the case where  $H$  is quadratic in  $X$ ,

$$C_{XX}(\omega) = \frac{2T\Gamma}{\omega^2\Gamma^2 + K^2} + \frac{2T\gamma v^2}{(\omega^2\gamma^2 + k^2)(\omega^2\Gamma^2 + K^2)} \quad (49)$$

and

$$\chi''_{XX}(\omega) = \frac{\Gamma\omega}{\omega^2\Gamma^2 + K^2} \quad (50)$$

so that

$$\omega C_{XX}(\omega) - 2T\chi''_{XX}(\omega) = 2T\omega \frac{\gamma v^2}{(\omega^2\gamma^2 + k^2)(\omega^2\Gamma^2 + K^2)}. \quad (51)$$

Of course

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{\omega\gamma_{11}}{T} [\omega C_{XX}(\omega) - 2T\chi''_{XX}(\omega)] = \frac{v^2}{\Gamma k + \gamma K} \quad (52)$$

which matches the entropy production we calculated earlier, but more important is the spectral decomposition of the entropy production:

$$\sigma_1^\omega = \frac{\omega^2\Gamma\gamma v^2}{\pi(\omega^2\gamma^2 + k^2)(\omega^2\Gamma^2 + K^2)}. \quad (53)$$

Note that  $\sigma_1^\omega$  is appreciable only at intermediate values of  $\omega$ , vanishing as  $\omega^2$  for  $\omega \rightarrow 0$  and as  $1/\omega^2$  as  $\omega \rightarrow \infty$ . The physical meaning of this distribution across frequencies merits some attention. Recall that  $\gamma/k$  is the correlation time of the coloured noise. For example, let  $\gamma/k \gg \Gamma/K$ , so that the correlations of the coloured noise are significant during the process of relaxation in the potential. By *high* frequency we mean  $\omega \gg K/\Gamma$ , which can be reinterpreted as probing timescales  $1/\omega$  on which you explore a mean-square distance  $T/\Gamma(1/\omega)$  smaller than the steady state positional variance  $T/K$ . That is, translational diffusion hasn't had a chance to act. It is this limit that our model is connected to that of [3]. By *low* frequency we mean  $\omega\gamma/k \ll 1$ , so one does not see the 'colour' of the 'noise'  $x$ , which then simply adds to the existing white noise in the inertia-less version of (8), just shifting the temperature of the overdamped oscillator coordinate  $X$ . Note that  $\sigma_1^\omega$  will display a broad plateau between the well-separated frequencies  $k/\gamma$  and  $K/\Gamma$ . Crucially from the point of view of experimental or numerical test of these ideas, the asymptotic vanishing of (53) at low and high frequencies implies that data that (a) sample  $X$  and  $x$  at time intervals much longer than  $\gamma/k$ , or (b) are limited to a duration much smaller than  $\Gamma/K$ , will give the impression of time-reversibility.

Our treatment so far takes all white noise strengths to be governed by a single bath temperature  $T$ . To turn our model into the traditional AOUP [3], we must allow

the temperature associated with the additive white noise in the equation for  $X$  to go to zero. The calculation presented above would then trivially give an infinite entropy production because backstepping of  $X$  is totally ruled out in the absence of the noise  $\eta$  in (8). One physically appealing way to reconcile this with the effective time-reversal invariance of the AOUP in a harmonic potential [3] is given in appendix E.

4.1.2. *HSR when the time-reversal operation involves a polarity flip.* In this case, we rewrite the equation for  $X$  as

$$\left(\dot{X} - \frac{v}{\Gamma}x\right) = -\frac{1}{\Gamma}\frac{\partial H}{\partial X} + \tilde{\eta} \quad (54)$$

where  $\langle \tilde{\eta}(t)\tilde{\eta}(t') \rangle = (2T/\Gamma)\delta(t-t')$ . Define  $(\dot{X} - (v/\Gamma)x) = V$ , the relative velocity. Let us define the response function  $\mathcal{R}_{VX} = \delta\langle V(t) \rangle / \delta h(t')$  where we have modified the Hamiltonian to  $H - hX$ . The change in action to first order in  $h$  is

$$\delta\mathcal{A} = -\frac{h}{2T} \int dt \left( \dot{X} - \frac{v}{\Gamma}x + \frac{1}{\Gamma}\frac{\partial H}{\partial X} \right). \quad (55)$$

Recalling the definition of  $V$  and using the definition of the response function, we immediately obtain

$$2T\mathcal{R}_{VX}(t, t') = \left\langle V(t) \left( V(t') + \frac{1}{\Gamma}\frac{\partial H}{\partial X}(t') \right) \right\rangle. \quad (56)$$

Note that  $V$  and  $X$  have opposite signs under time-reversal. Therefore, we now calculate the symmetric part of the response function

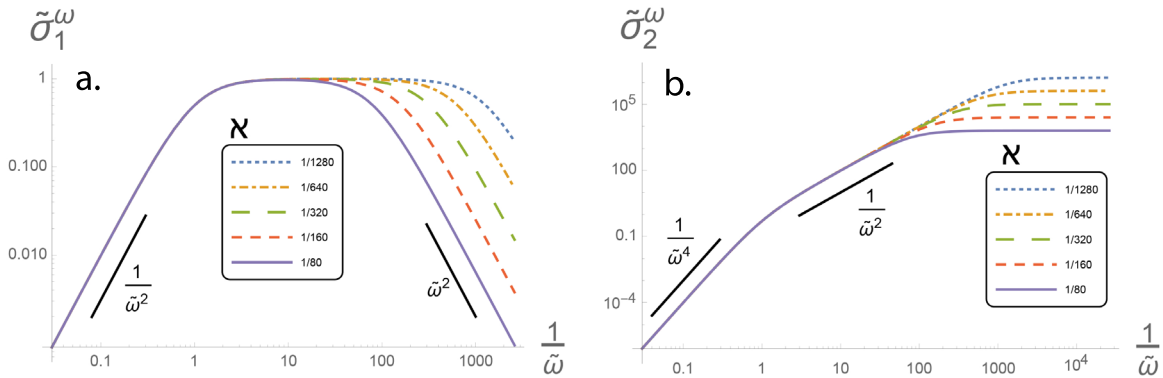
$$2T[\mathcal{R}_{VX}(t, t') + \mathcal{R}_{VX}(t', t)] = \langle V(t)V(t') \rangle + \langle V(t')V(t) \rangle + \frac{1}{\Gamma} \left[ \left\langle \dot{X}(t)\frac{\partial H}{\partial X}(t') \right\rangle + \left\langle \dot{X}(t')\frac{\partial H}{\partial X}(t) \right\rangle \right] - \frac{v}{\Gamma^2} \left[ \left\langle x(t)\frac{\partial H}{\partial X}(t') \right\rangle + \left\langle x(t')\frac{\partial H}{\partial X}(t) \right\rangle \right]. \quad (57)$$

If the system were in equilibrium, time-reversal symmetry would have resulted in the vanishing of the terms in square brackets on the right-hand side. This would have yielded the FDT  $T[\mathcal{R}_{VX}(t, t') + \mathcal{R}_{VX}(t', t)] = \langle V(t)V(t') \rangle$ . Out of equilibrium, the measure is not time-reversal symmetric. Nonetheless, for a quadratic potential,  $\langle \dot{X}(t')X(t) \rangle + \langle \dot{X}(t)X(t') \rangle$  would of course be 0. Even for a non-quadratic  $H$ , taking the limit  $t-t' \rightarrow 0$  and using time-translation invariance so that the correlation functions are only functions of  $t-t'$ ,

$$\lim_{t \rightarrow t'} \left\langle \dot{X}(t)\frac{\partial H}{\partial X}(t') \right\rangle = \langle \dot{H} \rangle = 0. \quad (58)$$

Defining  $\mathcal{C}(t, t') = \langle V(t)V(t') \rangle$ , we finally obtain the Harada–Sasa relation

$$-\frac{\Gamma}{T} \lim_{(t-t') \rightarrow 0} [T\{\mathcal{R}(t, t') + \mathcal{R}(t', t)\} - \mathcal{C}(t, t')] = \frac{v}{\Gamma T} \left\langle x(t)\frac{\partial H}{\partial X}(t) \right\rangle = \sigma_2. \quad (59)$$



**Figure 1.** (a) Frequency-dependent entropy production rate for a harmonically confined active polar dimer with translational diffusion, when time-reversal does not include a polarity-flip: as discussed, and consistent with [24], the rate vanishes both at small and large frequencies with a frequency-independent plateau in between which increases with decreasing  $\aleph$ . (b) When time-reversal includes a polarity-flip, the rate differs from (a) only by a factor of  $1/\tilde{\omega}^2$ .

Evaluating  $\mathcal{R}$  and  $\mathcal{C}$  in the frequency domain for the case where  $H$  is quadratic in  $X$  so that  $\partial_X H = KX$ , we obtain  $\sigma_2 = \int_{-\infty}^{\infty} \sigma_2^\omega d\omega$  with

$$\sigma_2^\omega = \frac{v^2 K^2 \gamma}{\pi \Gamma} \frac{1}{k^2 + \gamma^2 \omega^2} \frac{1}{K^2 + \Gamma^2 \omega^2} \tag{60}$$

whose zero-frequency limit is nonzero and independent of the stiffness of the harmonic potential. Note that the frequency-resolved entropy production rates in the two cases (53) and (60) have the same pole structures, and that their ratio is simply  $(\omega \Gamma / K)^2$ .

We now display the two different frequency-dependent entropy production rates for an AOUP in a harmonic potential in figure 1. For this, we non-dimensionalise the frequency by  $K/\Gamma$ , i.e.  $\tilde{\omega} = \omega \Gamma / K$ , and the entropy production rates as  $\tilde{\sigma}_{1,2}^\omega = \sigma_{1,2}^\omega \pi \gamma K^2 / v^2 \Gamma$ , with  $\tilde{\sigma}_1^\omega = \tilde{\omega}^2 \tilde{\sigma}_2^\omega = \tilde{\omega}^2 / \{(\tilde{\omega}^2 + 1)(\tilde{\omega}^2 + \aleph^2)\}$ , where the dimensionless constant  $\aleph = k \Gamma / K \gamma$ .

We have established that there are two natural equilibrium limits available for the AOUP depending upon whether the polarity is flipped or not as a part of the time-reversal operation. They can be directly measured from two distinct violations of fluctuation-dissipation relation. However, this choice is available only because  $x$  enters the equation for  $X$  only through the active term and since in the 0 activity limit the two equations decouple, one obtains two distinct equilibrium limits depending on whether  $x$  is odd or even under time-reversal. More technically, this choice is the result of the active term breaking both time-reversal and  $x \rightarrow -x$  symmetry in the model. If  $X$  and  $x$  were coupled through equilibrium terms in addition to the active one, such that the 0 activity limit did not possess this extra  $x \rightarrow -x$  symmetry, this would not be the case—the requirement that entropy production has to vanish in the absence of activity would fix the time-reversal characteristic of both  $x$  and  $X$ . We explicitly demonstrate this in appendix E.

J. Stat. Mech. (2018) 123201

4.1.3. *Entropy production of other active particle models.* We have also considered two other active particles: polar particles with a preferred speed and apolar particles. The entropy production for these two particles are

$$\sigma_P = \frac{\zeta \Delta \mu}{T} \left\langle \frac{P}{\Gamma(P^2)} \partial_X H \right\rangle, \quad (61)$$

where  $\Gamma(P^2) = \gamma_1 + \gamma_3 P^2$  and

$$\sigma_Q = -\frac{\zeta \Delta \mu}{\gamma T} \langle \mathbf{P} \cdot \mathbf{Q} \cdot \partial_{\mathbf{x}} H \rangle, \quad (62)$$

respectively. We have calculated the entropy production for the apolar particle in the overdamped limit.

## 4.2. Nonequilibrium dynamics of active field theories

4.2.1. *Entropy production for polar rods.* We first compute entropy production for polar rods on a substrate which we discussed in section 3.1.2. However, we will now consider a simplified and rescaled version of those equations which retains its essential active features:

$$\dot{\rho} = -v \Delta \mu \nabla \cdot (\rho \mathbf{p}) + \nabla^2 \frac{\delta H}{\delta \rho} + \xi \quad (63)$$

with  $\langle \xi(\mathbf{x}, t) \xi(\mathbf{x}', t') \rangle = -2T \nabla^2 \delta(\mathbf{x} - \mathbf{x}') \delta(t - t')$  and

$$\dot{\mathbf{p}} + \lambda \Delta \mu \mathbf{p} \cdot \nabla \mathbf{p} = -\frac{\delta H}{\delta \mathbf{p}} + \boldsymbol{\xi}^p \quad (64)$$

$\langle \boldsymbol{\xi}_p(\mathbf{x}, t) \boldsymbol{\xi}_p(\mathbf{x}', t') \rangle = 2T \mathbf{I} \delta(\mathbf{x} - \mathbf{x}') \delta(t - t')$  Here, we have ignored the noise cross-correlation between  $\rho$  and  $\mathbf{p}$  and retained the active advection of the density and polarisation at different speeds. The Onsager–Machlup action, ignoring the contribution from the noise discretisation, which does not contribute to the time-antisymmetric part of the action, is

$$\begin{aligned} \mathcal{A}^{OM} = & \frac{1}{4T} \int d\mathbf{x} dt \left[ - \left\{ \dot{\rho} + v \Delta \mu \nabla \cdot (\rho \mathbf{p}) - \nabla^2 \frac{\delta H}{\delta \rho} \right\} \nabla^{-2} \left\{ \dot{\rho} + v \Delta \mu \nabla \cdot (\rho \mathbf{p}) - \nabla^2 \frac{\delta H}{\delta \rho} \right\} \right. \\ & \left. + \left( \dot{\mathbf{p}} + \lambda \Delta \mu \mathbf{p} \cdot \nabla \mathbf{p} + \frac{\delta H}{\delta \mathbf{p}} \right)^2 \right]. \end{aligned} \quad (65)$$

Since both  $\mathcal{T}\rho = \rho$  and  $\mathcal{T}\mathbf{p} = \mathbf{p}$  the time antisymmetric part of this action is

$$\mathcal{S} = -\frac{\Delta \mu}{T} \int dt d\mathbf{x} \left[ -v \{ \dot{\rho} \nabla^{-2} \nabla \cdot (\rho \mathbf{p}) \} + \lambda \{ \dot{\mathbf{p}} \cdot (\mathbf{p} \cdot \nabla) \mathbf{p} \} \right]. \quad (66)$$

We show in appendix G that the entropy production rate  $\sigma = \lim_{t \rightarrow \infty} \mathcal{S}/t$  is related to the standard definitions of response and correlation function by a Harada–Sasa-like relation [3, 23]



$$\sigma = \frac{1}{T} \lim_{t \rightarrow 0} \int d\mathbf{x} \partial_t \left[ \nabla^{-2} (T[\mathcal{R}_{\rho\rho}(t) - \mathcal{R}_{\rho\rho}(-t)] + \partial_t C_{\rho\rho}) - (T[\mathcal{R}_{pp}(t) - \mathcal{R}_{pp}(-t)] + \partial_t C_{pp}) \right] \quad (67)$$

where  $\mathcal{R}_{ij}$  is the response of the field  $j$  to a perturbation in the field  $i$  and  $C_{ij}$  is the correlation of the fields  $i$  and  $j$ . We also show that in contrast to active model B [64], the entropy production does not vanish in the  $T \rightarrow 0$  limit even in a homogeneously polarised phase in appendix H.

*4.2.2. Entropy production for the model with velocity-alignment.* We will now calculate the entropy-production of the active variant of the Navier–Stokes equation we discussed in section 3.2.3. For simplicity, let us set the viscosity  $\eta = 0$ . The action is

$$\mathcal{A} = \frac{1}{4T} \int dt d\mathbf{x} \frac{1}{\Gamma_{gg}(v)} \left[ \partial_t \mathbf{g} + \nabla \cdot (\mathbf{g}\mathbf{v}) + \rho \nabla \frac{\delta H}{\delta \rho} + \mathbf{\Gamma}_{gg} \mathbf{v} - \Delta \mu \zeta \mathbf{v} \right]^2 \quad (68)$$

since  $\mathcal{T}\mathbf{g} = -\mathbf{g}$ , the antisymmetric part of this action, which can not be expressed as a total derivative, is

$$\mathcal{S} = \frac{\Delta \mu}{T} \int dt d\mathbf{x} \frac{\zeta}{\Gamma_{gg}(v)} \mathbf{v} \cdot \left[ \partial_t \mathbf{g} + \nabla \cdot (\mathbf{g}\mathbf{v}) + \rho \nabla \frac{\delta H}{\delta \rho} \right]. \quad (69)$$

Note that there are two crucial requirements for non-zero entropy production: i. the presence of  $\Delta \mu$  and ii. the multiplicativity of the noise. If the noise were not multiplicative, (69) would have been a total derivative which would not lead to any entropy production. Of course, in that case the cubic velocity nonlinearity required for saturating the speed when the linear damping is negative, would also have to be an active term and that would have led to a non-zero entropy.

## 5. Conclusion

In this paper we first presented in some detail the stochastic extension of the nonequilibrium thermodynamic construction [4–8] of the dynamical equations for active systems, i.e. systems in which detailed balance is broken homogeneously at the level of the constituent degrees of freedom. Through this treatment we saw that all active systems discussed in the literature can be obtained from a corresponding equilibrium system in which the force on an internal coordinate, identified as a chemical degree of freedom, is maintained at a nonzero constant value, while all degrees of freedom are in contact with a *common* thermal bath. This includes apparently ‘two-temperature’ models, such as apolar rods which we discuss as well as those studied in [76], and systems with non-mutual pair interactions [77–79]. For active polar particles we then use the ratio of the experimentally measurable large deviation functions for trajectories in position-polarisation space to demonstrate that there are two *a priori* equally valid answers to the question how far a system is from equilibrium. Our approach requires some comment: active systems are in general complicated and quantitative measurements of the total entropy production rate, which is related to the total heat dissipation rate, are generally impractical to carry out. What is desired instead is that, given a measurement

of certain quantities, to understand how different it is from a measurement of those quantities if the system were in equilibrium. This limited question is answered by the ‘entropy production’ we calculate though we show that even this may not have unique answers in certain cases. Our work is similar in spirit to that of [80, 81], who try to find, in a configuration space spanned by experimentally measured quantities, a circulating dissipative current whose presence signifies the breaking of detailed balance. As long as one chooses two variables with the same sign under time-reversal, measuring the circulating current provides a direct quantitative measure of a system’s distance from equilibrium, and can be easily related to the measure we discuss [82]. We too take an objective, empirical approach to the extent of time-reversal breaking based on measurements on a subset of degrees of freedom, with no implied link to heat. Rather than a vectorial quantity such as a configuration-space current, we work with a scalar measure.

Finally, turning to field theories of polar active systems, we have clarified the two equilibrium limits of the Toner–Tu equation—polar liquid crystal and the Navier–Stokes equation—and explained the origin and the consequence of nonequilibrium driving in both of them. Our work here clarifies a long-standing confusion about the self-advective nonlinearity of the Toner–Tu equation, which we show requires breaking of both detailed balance and Galilean invariance.

While our work in this paper has been mainly clarificatory, we have made testable predictions regarding entropy production rate of polar particles. In particular, we calculated the frequency-dependence of the entropy production in harmonic traps and demonstrated that when only the position of the particle is measured, there is no way to distinguish the dynamics from an equilibrium one. We look forward to tests of our predictions on model active systems in the colloidal or granular domain.

## Acknowledgments

We thank Mike Cates, Suraj Shankar and M Cristina Marchetti for illuminating discussions. LPD thanks the Department of Physics, Indian Institute of Science, for hospitality and support. SR was supported by a JC Bose Fellowship of the SERB (India) and the Tata Education and Development Trust.

## Appendix A. Active dynamics through the coupling of physical and chemical velocities

In this appendix, we follow [26] and retain an extra chemical velocity variable  $\Phi$ , which is conjugate to the chemical coordinate  $n$ . We obtain active equations of motion for each system described in the main text without invoking any direct coupling between  $n$  and other variables, but through an entrainment of the physical velocity by the chemical velocity or an antisymmetric reactive coupling between the chemical and the physical velocities. We take the dynamics of  $n$  to be  $\partial_t n = \Phi$ . Our discussion highlights the fact that multiple thermodynamically consistent dynamical models lead to the same effective active equations of motion. While active equations of motion for the models discussed in section 3.1 have already been derived using this approach in [26], in this

appendix we focus on the models discussed in section 3.2. Unlike the models discussed in [26], in which the coupling between the physical and chemical velocities were dissipative in nature, these models naturally require a *reversible* coupling between physical and chemical velocities.

### A.1. Reversible coupling of physical and chemical velocities

We now derive the equations of motion for (i) an active polar model with inertia and a negative linear damping which can be viewed as an active extension of the Drude model, (ii) an apolar variant of the AOUP and (iii) a field theory of a polar model with inertia and negative linear damping.

*A.1.1. Active particle with preferred speed.* Start with the system of equations

$$\dot{X} = \partial_P H; \dot{n} = \partial_\Phi H \quad (\text{A.1})$$

$$\dot{P} = -\Gamma_{11}\partial_P H + \Gamma_{12}\partial_\Phi H - \partial_X H + \xi^P = -\Gamma_{11}\partial_P H + \zeta P\partial_\Phi H - \partial_X H + \xi^P, \quad (\text{A.2})$$

$$\dot{\Phi} = -\Gamma_{22}\partial_\Phi H - \Gamma_{12}\partial_P H + T\partial_P\Gamma_{12} - \partial_n H + \xi^\Phi = -\Gamma_{22}\partial_\Phi H - \zeta P\partial_P H + T\zeta - \partial_n H + \xi^\Phi \quad (\text{A.3})$$

with  $\langle \xi^P(t)\xi^P(0) \rangle = 2\Gamma_{11}T\delta(t)$  and  $\langle \xi^\Phi(t)\xi^\Phi(0) \rangle = 2\Gamma_{22}T\delta(t)$ . Here, we have coupled  $P$  and  $\Phi$  reversibly through a Poisson bracket  $[P, \Phi] = \Gamma_{12} = \zeta P$ . The term  $T\partial_P\Gamma_{12} = T\zeta$  arises routinely in the projection-operator construction of generalized Langevin equations [30, 34], and its presence is formally required to ensure that the Gibbs–Boltzmann distribution is the stationary solution to the corresponding Fokker–Planck equation. In most other systems this term conveniently vanishes; here it does not.

Ignoring inertia in (A.3), we obtain

$$\partial_\Phi H = -\frac{\zeta P}{\Gamma_{22}}\partial_P H + \frac{T\zeta}{\Gamma_{22}} - \frac{1}{\Gamma_{22}}\frac{\partial H}{\partial n} + \frac{1}{\Gamma_{22}}\xi^\Phi \quad (\text{A.4})$$

which allows us to eliminate  $\partial_\Phi H$  from (A.2), yielding, for the case where  $H$  is quadratic in  $P$  with  $\partial_P H = P$ ,

$$\dot{P} = -\left(\Gamma_{11} + \frac{\zeta}{\Gamma_{22}}\frac{\partial H}{\partial n} - \frac{T\zeta^2}{\Gamma_{22}}\right)P - \frac{\zeta^2 P^2}{\Gamma_{22}} - \frac{\partial H}{\partial X} + \eta \quad (\text{A.5})$$

where

$$\eta \equiv (\zeta P/\Gamma_{22})\xi^\Phi + \xi^P, \langle \eta(t)\eta(0) \rangle = 2T\left(\frac{\zeta^2 P^2}{\Gamma_{22}} + \Gamma_{11}\right)\delta(t), \quad (\text{A.6})$$

is a multiplicative noise in the  $P$  equation, with the Stratonovich interpretation, as can be seen by carrying out the adiabatic elimination in detail à la [37, 83]. Alternatively one can check this ex post facto by confirming that the corresponding Fokker–Planck equation has the correct equilibrium solution. The term  $-(T\zeta^2/\Gamma_{22})P$  in (A.5), whose origin is the derivative of the  $[P, \Phi]$  Poisson bracket, must now be understood as a noise-induced drift. Now, we introduce activity by holding  $\partial_n H = \Delta\mu$  constant. This leads to the equation of motion for  $P$

$$\dot{P} = - \left( \Gamma_{11} + \frac{\zeta}{\Gamma_{22}} \Delta\mu - \frac{T\zeta^2}{\Gamma_{22}} + \frac{\zeta^2 P^2}{\Gamma_{22}} \right) P - \frac{\partial H}{\partial X} + \eta. \quad (\text{A.7})$$

Note that the  $P$ -independent part of the friction is now liberated from the noise strength because of the  $\Delta\mu$ -dependent term, which can even make it negative. Despite superficial appearances, the noise-induced drift  $-(T\zeta^2/\Gamma_{22})P$  has no such effect, and arises simply as a consequence of the multiplicative noise [31]. In an anti-Itô [31] description, the noise-induced drift vanishes. In all interpretations, of course, the probability distribution of  $P$  is Maxwellian if  $\Delta\mu = 0$ , and the particle is not self-propelled. If  $\zeta < -\Gamma_{11}\Gamma_{22}/\Delta\mu$ , the zero-velocity state is on average unstable, and the particle moves with preferred velocity  $\pm\sqrt{(|\zeta|\Delta\mu - \Gamma_{11}\Gamma_{22})/\zeta^2}$  in the limit of low noise. We have thus succeeded in producing a model self-propelled particle with inertia within this consistent stochastic framework.

*A.1.2. Apolar AOUP.* Take the coordinate-like variables to be  $(\mathbf{X}, n, \mathbf{Q})$  where  $\mathbf{Q}$  is a traceless symmetric second-rank tensor. We assume that  $\mathbf{Q}$  is autonomous and obeys a purely relaxational stochastic dynamics:

$$\Gamma_{33} \dot{\mathbf{Q}} = -\mathbf{Q} + \xi^Q \quad (\text{A.8})$$

with  $\langle \xi^Q(t) \xi^Q(t') \rangle = 2\Gamma_{33} T \delta(t - t')$ . As before, the momenta conjugate to  $(\mathbf{X}, n)$  are  $(\mathbf{P}, \Phi)$ . Take the system of equations

$$\dot{\mathbf{X}} = \mathbf{P} \quad (\text{A.9})$$

$$\dot{\mathbf{P}} = -\Gamma_{11} \mathbf{P} + \Gamma_{12} \Phi - \frac{\partial H}{\partial \mathbf{X}} + \boldsymbol{\xi}^P = -\Gamma_{11} \mathbf{P} + \zeta \mathbf{Q} \cdot \mathbf{P} \Phi - \frac{\partial H}{\partial \mathbf{X}} + \boldsymbol{\xi}^P \quad (\text{A.10})$$

with  $\langle \boldsymbol{\xi}^P(t) \boldsymbol{\xi}^P(t') \rangle = 2\Gamma_{11} T \mathbf{I} \delta(t - t')$ , where  $\mathbf{I}$  is the identity tensor.

$$\dot{n} = \Phi \quad (\text{A.11})$$

$$\dot{\Phi} = -\Gamma_{22} \Phi - \Gamma_{12} \cdot \mathbf{P} - \frac{\partial H}{\partial n} + \xi^\Phi = -\Gamma_{22} \Phi - \zeta \mathbf{P} \cdot \mathbf{Q} \cdot \mathbf{P} - \frac{\partial H}{\partial n} + \xi^\Phi. \quad (\text{A.12})$$

Note that since  $\mathbf{Q}$  is traceless,  $\nabla_{\mathbf{P}} \cdot (\mathbf{Q} \cdot \mathbf{P})$  is 0 and therefore, the derivative of the Poisson bracket term is 0. Again overdamping (A.12) and replacing  $\Phi$  in (A.10), we obtain

$$\begin{aligned} \dot{\mathbf{P}} &= - \left[ \Gamma_{11} \mathbf{I} + \frac{\zeta}{\Gamma_{22}} \frac{\partial H}{\partial n} \mathbf{Q} + \frac{\zeta^2}{\Gamma_{22}} (\mathbf{Q} \cdot \mathbf{P})(\mathbf{Q} \cdot \mathbf{P}) \right] \cdot \mathbf{P} - \frac{\partial H}{\partial \mathbf{X}} + \frac{\zeta}{\Gamma_{22}} \mathbf{Q} \cdot \mathbf{P} \xi^\Phi + \boldsymbol{\xi}^P \\ &= - \left[ \Gamma_{11} \mathbf{I} + \frac{\zeta}{\Gamma_{22}} \frac{\partial H}{\partial n} \mathbf{Q} + \frac{\zeta^2}{\Gamma_{22}} (\mathbf{Q} \cdot \mathbf{P})(\mathbf{Q} \cdot \mathbf{P}) \right] \cdot \mathbf{P} - \frac{\partial H}{\partial \mathbf{X}} + \boldsymbol{\eta} \end{aligned} \quad (\text{A.13})$$

with the noise correlator  $\langle \boldsymbol{\eta}(t) \boldsymbol{\eta}(t') \rangle = 2T [\Gamma_{11} \mathbf{I} + (\zeta^2/\Gamma_{22})(\mathbf{Q} \cdot \mathbf{P})(\mathbf{Q} \cdot \mathbf{P})]$ . This equation should again be interpreted using the Stratonovich interpretation. However, due to tracelessness of  $\mathbf{Q}$ , no interpretation-dependent noise is required to transfer between interpretations. In other words, this is a multiplicative noise equation in which there

is no possibility of noise-induced drift. Now holding  $\partial_n H = \Delta\mu$  constant, we obtain a nonequilibrium  $\mathbf{Q}$ -dependent damping:

$$\dot{\mathbf{P}} = - \left[ \Gamma_{11} \mathbf{l} + \frac{\zeta \Delta\mu}{\Gamma_{22}} \mathbf{Q} + \frac{\zeta^2}{\Gamma_{22}} (\mathbf{Q} \cdot \mathbf{P})(\mathbf{Q} \cdot \mathbf{P}) \right] \cdot \mathbf{P} - \frac{\partial H}{\partial \mathbf{X}} + \boldsymbol{\eta}. \quad (\text{A.14})$$

If we now take the damping of  $\mathbf{P}$  in (A.10) to depend on  $\mathbf{Q}$  instead of being isotropic, we can cancel the passive  $\mathbf{Q}$  dependent part of (A.14). Write the damping in (A.10) as  $-\Gamma_P \cdot \mathbf{P} = -(\Gamma_{11} \mathbf{l} - \zeta^2 (\mathbf{Q} \cdot \mathbf{P})(\mathbf{Q} \cdot \mathbf{P}) / \Gamma_{22}) \cdot \mathbf{P}$  which yields the simplest apolar version of AOUP:

$$\dot{\mathbf{P}} = - \left[ \Gamma_{11} \mathbf{l} + \frac{\zeta \Delta\mu}{\Gamma_{22}} \mathbf{Q} \right] \cdot \mathbf{P} - \frac{\partial H}{\partial \mathbf{X}} + \boldsymbol{\eta}_1 \quad (\text{A.15})$$

with  $\langle \boldsymbol{\eta}_1(t) \boldsymbol{\eta}_1(t') \rangle = 2T \Gamma_{11} \mathbf{l}$ . Taking the overdamped limit is now trivial.

*A.1.3. Active field theory with velocity alignment.* We now start with the Navier–Stokes equation with friction.

$$\partial_t \rho = -\nabla \cdot \mathbf{g} \quad (\text{A.16})$$

$$\partial_t \mathbf{g} + \nabla \cdot (\mathbf{v} \mathbf{g}) = -\rho \nabla \frac{\delta F}{\delta \rho} - \Gamma \mathbf{v} + \eta \nabla^2 \mathbf{v} + \boldsymbol{\xi}^g \quad (\text{A.17})$$

with  $\langle \boldsymbol{\xi}^g(\mathbf{x}, t) \boldsymbol{\xi}^g(\mathbf{x}', t') \rangle = 2T \mathbf{l} (\Gamma - \eta \nabla^2) \delta(\mathbf{x} - \mathbf{x}') \delta(t - t')$ , and  $H_0 = \int_x \mathbf{g}^2 / 2\rho + F[\rho]$ . We then introduce the interaction with the chemical field by extending the hamiltonian  $H = H_0 + n \Delta\mu$  and we consider the reactive coupling

$$\{\Phi(\mathbf{x}'), \mathbf{g}(\mathbf{x})\} \frac{\delta H}{\delta \Phi(\mathbf{x}')},$$

between  $\Phi$  and  $\mathbf{g}$  in the momentum equation. To the lowest order,  $\{\Phi(\mathbf{x}'), \mathbf{g}(\mathbf{x})\} = \alpha_1(\rho) \mathbf{v}(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}')$ . Note that the derivative of the Poisson bracket with  $\mathbf{g}$  is not zero:  $\delta\{\Phi(\mathbf{x}'), \mathbf{g}(\mathbf{x})\} / \delta \mathbf{g}(\mathbf{x}') \neq 0$ .

We now write the dynamical equation for  $\Phi$ :

$$\partial_t \Phi = -\gamma \Phi - \frac{\delta F}{\delta n} + \int d\mathbf{x}' \left[ \{\mathbf{g}(\mathbf{x}'), \Phi(\mathbf{x})\} \frac{\delta H}{\delta \mathbf{g}(\mathbf{x}')} + T \frac{\delta\{\Phi(\mathbf{x}'), \mathbf{g}(\mathbf{x})\}}{\delta \mathbf{g}(\mathbf{x}')} \right] + \xi^\Phi \quad (\text{A.18})$$

where the noise correlation is  $\langle \xi^\Phi(\mathbf{x}, t), \xi^\Phi(\mathbf{x}', t') \rangle = 2T \gamma \delta(\mathbf{x} - \mathbf{x}') \delta(t - t')$ . The finite-temperature contribution to the reactive piece is evaluated as

$$T \frac{\delta\{\Phi(\mathbf{x}'), \mathbf{g}(\mathbf{x})\}}{\delta \mathbf{g}(\mathbf{x}')} = T \alpha_1(\rho) \frac{1}{\rho(\mathbf{x})} \delta(\mathbf{x} - \mathbf{x}') \quad (\text{A.19})$$

while

$$\{\mathbf{g}(\mathbf{x}'), \Phi(\mathbf{x})\} \frac{\delta H}{\delta \mathbf{g}(\mathbf{x}')} = -\alpha_1(\rho) \mathbf{v}(\mathbf{x}) \cdot \mathbf{v}(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}'). \quad (\text{A.20})$$

Finally, taking the overdamped limit in the  $\Phi$  equation, we can solve for  $\Phi$  as

$$\Phi = -\frac{1}{\gamma} \frac{\delta H}{\delta n} - \frac{\alpha_1(\rho)}{\gamma} \mathbf{v}^2 + T \frac{\alpha_1(\rho)}{\rho\gamma} + \frac{1}{\gamma} \xi^\Phi \quad (\text{A.21})$$

$$\int d\mathbf{x}' \{ \Phi(\mathbf{x}'), \mathbf{g}(\mathbf{x}) \} \frac{\delta H}{\delta \Phi(\mathbf{x}')} = \alpha_1(\rho) \mathbf{v} \Phi = - \left( \frac{\alpha_1(\rho)}{\gamma} \frac{\delta H}{\delta n} - T \frac{\alpha_1^2(\rho)}{\rho\gamma} \right) \mathbf{v} - \frac{\alpha_1^2(\rho)}{\gamma} \mathbf{v}^2 \mathbf{v} + \frac{\alpha_1(\rho)}{\gamma} \mathbf{v} \xi^\Phi. \quad (\text{A.22})$$

Finally, noting that  $\delta H/\delta n = \Delta\mu$  is the constant chemical drive, and using (24) we have the Toner–Tu equation:

$$\rho(\partial_t \mathbf{v} + \mathbf{v} \cdot \nabla \cdot \mathbf{v}) = -\rho \nabla \frac{\delta F}{\delta \rho} - \left( \Gamma + \frac{\alpha_1(\rho)}{\gamma} \Delta\mu - T \frac{\alpha_1^2(\rho)}{\rho\gamma} \right) \mathbf{v} - \frac{\alpha_1^2(\rho)}{\gamma} \mathbf{v}^2 \mathbf{v} + \eta \nabla^2 \mathbf{v} + \left( \xi^g + \frac{\alpha_1(\rho)}{\gamma} \mathbf{v} \xi^\Phi \right). \quad (\text{A.23})$$

The noise is again multiplicative with the correlator

$$\begin{aligned} \left\langle \left( \xi^g + \frac{\alpha_1(\rho)}{\gamma} \mathbf{v} \xi^\Phi \right) (\mathbf{x}, t) \left( \xi^g + \frac{\alpha_1(\rho)}{\gamma} \mathbf{v} \xi^\Phi \right) (\mathbf{x}', t') \right\rangle &= \langle \xi^v(\mathbf{x}, t) \xi^v(\mathbf{x}', t') \rangle \\ &= 2T [\mathbf{I}(\Gamma - \eta \nabla^2) + \alpha_1(\rho)^2 \mathbf{v} \mathbf{v} / \gamma] \delta(\mathbf{x} - \mathbf{x}') \delta(t - t'). \end{aligned} \quad (\text{A.24})$$

This equation should also be interpreted in the Stratonovich interpretation, since it should have a steady state  $\exp(-H_0/T)$  when  $\Delta\mu = 0$ .

## Appendix B. Activity through fixed reaction velocity

In the main text, we introduced activity through a fixed chemical potential difference between reactants and products, i.e.  $\partial H/\partial n = \text{const.}$ . This is in line with the traditional rationalisation for active matter equations where the chemical potential difference between a fuel (for instance, ATP of actomyosin networks in cells and sugar in bacterial systems) and its reaction by-products (for instance, ADP in actomyosin systems) is assumed to be constant. However, this choice is not essential—indeed, one can imagine bacteria or birds or even artificial particles which have fixed reaction velocity and adjusts  $\partial H/\partial n$  to hold that constant (i.e. uses as much or as little fuel as required to perform a fixed number of duty cycles in a second). We now demonstrate that an equation of the form (8) can be derived even by holding  $\Phi$  constant—in other words, similar effective equations result irrespective of whether microscopically a rate ( $\Phi$ ) or a field ( $\partial H/\partial n$ ) is held constant. As in section 2, take the model (with  $\mathcal{Q} = (X, n, x)$  and  $\mathcal{P} = (P, \Phi, p)$ )

$$\dot{\mathcal{Q}} = \frac{\partial H}{\partial \mathcal{Q}} \quad \dot{\mathcal{P}} = -\frac{\partial H}{\partial \mathcal{P}} - \Gamma \cdot \frac{\partial H}{\partial \mathcal{P}} + \xi \quad (\text{B.1})$$

where the noise  $\xi = (\xi^P, \xi^\Phi, \xi^p)$  has a correlation  $\langle \xi(t) \xi(t') \rangle = 2T \Gamma^S \delta(t - t')$ . and the friction matrix is

$$\Gamma = \begin{pmatrix} \Gamma & -\gamma_{12}x & 0 \\ -\gamma_{12}x & \gamma_{22} & 0 \\ 0 & 0 & \gamma \end{pmatrix}. \quad (\text{B.2})$$

After taking the overdamped limit, we now take  $\Phi$  to be held constant instead of  $\partial H/\partial n$ . This leads to the equations of motion

$$\Gamma \dot{X} = -\frac{\partial H}{\partial X} + \Phi \gamma_{12} x + \xi^P \tag{B.3}$$

$$\gamma \dot{x} = -kx + \xi^P \tag{B.4}$$

where we have assumed that  $H$  is quadratic in  $x$  (i.e.  $\partial_x H = kx$ ). This has the same form as the overdamped version of (8) and (9) except for the coefficient of the active coupling.

### Appendix C. Inertial active particles

In the main text, we have only considered the dynamics of overdamped active particles. In this appendix, we briefly consider the underdamped case. The explicit underdamped versions of the equations of motion (8) and (9) are

$$\ddot{X} + \Gamma \dot{X} + \frac{\partial H}{\partial X} - vx = \eta \tag{C.1}$$

$$\gamma \dot{x} + kx = \xi^P \tag{C.2}$$

with the noise correlations  $\langle \eta(t)\eta(t') \rangle = 2T\Gamma\delta(t-t')$  and  $\langle \xi^P(t)\xi^P(t') \rangle = 2T\gamma\delta(t-t')$ . The action is

$$\mathcal{A} = \frac{1}{4T} \int dt \frac{1}{\Gamma} (\ddot{X} + \Gamma \dot{X} + \frac{\partial H}{\partial X} - vx)^2 + \frac{1}{\gamma} (\gamma \dot{x} + kx)^2. \tag{C.3}$$

Reversing  $t \rightarrow -t$  while holding  $x$  fixed, we calculate the KL divergence which is

$$\mathcal{S}_1 = \frac{v}{T} \langle \dot{X} x \rangle \tag{C.4}$$

where we have as usual assumed ergodicity and replaced a time-average with an ensemble average. If we measure the KL divergence conditioned on  $x \rightarrow -x$  under time-reversal, we obtain

$$\mathcal{S}_2 = \frac{v}{\Gamma T} \langle x(\ddot{X} + \partial_X H) \rangle. \tag{C.5}$$

The sum  $\mathcal{S}_1 + \mathcal{S}_2 = v^2 \langle x^2 \rangle / T\Gamma = v^2 / \Gamma k$  as in the case without inertia.

We now show that our conclusion that a nonvanishing entropy production for a particle in a harmonic trap is only possible when the dynamics of the auxiliary variable  $x$  is explicitly tracked is valid even in the underdamped limit. In this case, (C.1) is modified to

$$\ddot{X} + \Gamma \dot{X} + \partial_X H = \eta. \tag{C.6}$$

The action case for a general noise is

$$\mathcal{A} = \frac{1}{4T} \int dt \left[ \nu_1 \left( \ddot{X} + \Gamma \dot{X} + \partial_X H \right)^2 + \nu_2 \gamma^2 \left\{ \partial_t \left( \ddot{X} + \Gamma \dot{X} + \partial_X H \right) \right\}^2 + \nu_3 \gamma^4 \left\{ \partial_t^2 \left( \ddot{X} + \Gamma \dot{X} + \partial_X H \right) \right\}^2 + \dots \right] \tag{C.7}$$



using the earlier definitions. The time-asymmetric part of this can be trivially calculated (integrating by parts repeatedly)

$$\mathcal{S} = \frac{1}{T} \int dt \left[ -\nu_1(\dot{X})(\ddot{X} + \partial_X H) + \nu_2(\partial_t^3 X)(\ddot{X} + \partial_X H) - \nu_3(\partial_t^5 X)(\ddot{X} + \partial_X H) + \dots \right]. \quad (\text{C.8})$$

Thus, inertia only contributes terms  $(\partial_t^{2n+3} X)X$ . This is just  $(1/2)\partial_t(\partial_t^{n+1} X)^2$  (integrating by parts  $n + 1$  times). All these terms vanish irrespective of the potential. We have already shown that the terms with  $\partial_X H$  vanish if  $H$  is harmonic in  $X$ . Therefore, the presence of inertia has no effect on the entropy production for any noise when one only records the particle position.

### Appendix D. Entropy production due to particle driven by pure coloured noise

In this appendix, we calculate the entropy production rate of a particle in a harmonic potential and driven by a monochromatic coloured noise using (36) and demonstrate that the result is equivalent to [3]. The equation of motion of such a particle is  $\Gamma\dot{X} + \partial_x H = v\xi$  with the noise correlation

$$\langle \xi_\omega \xi_{-\omega} \rangle = \frac{2T\gamma}{1 + \gamma^2\omega^2}, \quad (\text{D.1})$$

and therefore,

$$\mathcal{G}_\omega = \frac{1 + \gamma^2\omega^2}{v^2\gamma}. \quad (\text{D.2})$$

This leads to the action

$$\mathcal{A} = \frac{1}{4Tv^2\gamma} \int d\omega \eta_{-\omega} (1 + \gamma^2\omega^2) \eta_\omega = \frac{1}{4Tv^2\gamma} \int dt (\eta(t)^2 - \gamma^2 \dot{\eta}(t)\eta(t)) = \frac{1}{4Tv^2\gamma} \int dt [\eta(t)^2 + \gamma^2(\dot{\eta}(t))^2] \quad (\text{D.3})$$

which upon transformation to the  $X$  variables is

$$\mathcal{A} = \frac{1}{4Tv^2\gamma} \int dt [(\Gamma\dot{X} + \partial_X H)^2 + \gamma^2(\Gamma\ddot{X} + \dot{X}\partial_X^2 H)^2]. \quad (\text{D.4})$$

Adding only total derivative terms, this can be written as

$$\mathcal{A} = \frac{1}{4Tv^2\gamma} \int dt [\gamma\Gamma\ddot{X} + (\Gamma + \gamma\partial_X^2 H)\dot{X} + \partial_X H]^2. \quad (\text{D.5})$$

The time-reversed action is

$$\mathcal{A}^R = \frac{1}{4Tv^2\gamma} \int dt [\gamma\Gamma\ddot{X} - (\Gamma + \gamma\partial_X^2 H)\dot{X} + \partial_X H]^2 \quad (\text{D.6})$$

and the path-dependent antisymmetric part of this action is

$$\mathcal{S} = \frac{\Gamma\gamma}{2Tv^2} \int dt \dot{X}^3 \partial_X^3 H. \quad (\text{D.7})$$

This is the result for the KL divergence obtained in [3].

### Appendix E. Entropy production and time-reversal in general systems with two degrees of freedom

We show that two distinct equilibrium limits are possible for AOUP/ ABP only because the active term—in addition to breaking time-reversal symmetry—is also the only coupling between  $X$  and  $x$ . In other words, activity in this case breaks two different symmetries—(a) time-reversal and (b) independent  $X \rightarrow -X$  and  $x \rightarrow -x$  symmetries. When the active term is present, the only symmetry is under joint  $X \rightarrow -X$  and  $x \rightarrow -x$ . (if the active term is proportional to  $x^2$  instead of  $x$ , thus not breaking this symmetry, there is no difference between entropy production irrespective of whether  $x \rightarrow x$  or  $-x$  under  $t \rightarrow -t$ ). If the passive system does not have independent  $X \rightarrow -X$  and  $x \rightarrow -x$  symmetry, only one of the two measures vanish with  $\Delta\mu$ . (Strictly, the important symmetry is not under  $X \rightarrow -X$  but under  $x \rightarrow -x$ . That is the symmetry whose presence in the passive limit of ABP/ AOUP and absence in the active case leads to two distinct equilibrium limits, when polarity is flipped and when it is not). For example, if an ABP Hamiltonian contained a coupling  $\mathbf{p} \cdot \mathbf{r}$  where  $\mathbf{p}$  is the polarisation and  $\mathbf{r}$  is the position, then the ‘entropy production rate’ calculated by flipping  $\mathbf{p}$  will not vanish in the  $\Delta\mu \rightarrow 0$  limit but the one calculated without flipping  $\mathbf{p}$  will.

Take the phenomenological equations

$$\dot{X} = -\Gamma \frac{\partial H}{\partial X} - \gamma_{12} \frac{\partial H}{\partial x} + \zeta \Delta\mu x + \xi_X \tag{E.1}$$

$$\dot{x} = -\gamma_{22} \frac{\partial H}{\partial x} - \gamma_{12} \frac{\partial H}{\partial X} + \xi_x \tag{E.2}$$

with the noise correlators  $\langle \xi_X(t) \xi_X(t') \rangle = 2T\Gamma\delta(t-t')$ ,  $\langle \xi_x(t) \xi_x(t') \rangle = 2T\gamma_{22}\delta(t-t')$  and  $\langle \xi_X(t) \xi_x(t') \rangle = 2T\gamma_{12}\delta(t-t')$ . This has the action

$$\begin{aligned} \mathcal{A} = & \frac{1}{4T(\Gamma\gamma_{22} - \gamma_{12}^2)} \int dt \gamma_{22} \left[ \dot{X} + \Gamma \frac{\partial H}{\partial X} + \gamma_{12} \frac{\partial H}{\partial x} - \zeta \Delta\mu x \right]^2 + \Gamma \left[ \dot{x} + \gamma_{22} \frac{\partial H}{\partial x} + \gamma_{12} \frac{\partial H}{\partial X} \right]^2 \\ & - 2\gamma_{12} \left[ \dot{X} + \Gamma \frac{\partial H}{\partial X} + \gamma_{12} \frac{\partial H}{\partial x} - \zeta \Delta\mu x \right] \left[ \dot{x} + \gamma_{22} \frac{\partial H}{\partial x} + \gamma_{12} \frac{\partial H}{\partial X} \right]. \end{aligned} \tag{E.3}$$

Implement two different time reversal protocols (i)  $t \rightarrow -t$ ,  $X \rightarrow X$ ,  $x \rightarrow x$  (ii)  $t \rightarrow -t$ ,  $X \rightarrow X$ ,  $x \rightarrow -x$ . Note that in case (ii) in the absence of  $\Delta\mu$  the off-diagonal kinetic coefficient in the  $x$  equation is reversible and should have the opposite sign in the equilibrium limit. For (i), the time-reversed action is

$$\begin{aligned} A_{(i)}^R = & \frac{1}{4T(\Gamma\gamma_{22} - \gamma_{12}^2)} \int dt \gamma_{22} \left[ -\dot{X} + \Gamma \frac{\partial H}{\partial X} + \gamma_{12} \frac{\partial H}{\partial x} - \zeta \Delta\mu x \right]^2 + \Gamma \left[ -\dot{x} + \gamma_{22} \frac{\partial H}{\partial x} + \gamma_{12} \frac{\partial H}{\partial X} \right]^2 \\ & - 2\gamma_{12} \left[ -\dot{X} + \Gamma \frac{\partial H}{\partial X} + \gamma_{12} \frac{\partial H}{\partial x} - \zeta \Delta\mu x \right] \left[ -\dot{x} + \gamma_{22} \frac{\partial H}{\partial x} + \gamma_{12} \frac{\partial H}{\partial X} \right]. \end{aligned} \tag{E.4}$$

Defining  $\sigma_{(i)} = \lim_{t \rightarrow \infty} (\mathcal{A}_{(i)}^R - \mathcal{A})/t$ , we get

$$\begin{aligned} \sigma_{(i)} &= \lim_{t \rightarrow \infty} \frac{1}{t} \frac{1}{T} \int dt \left[ - \left( \dot{X} \frac{\partial H}{\partial X} + \dot{x} \frac{\partial H}{\partial x} \right) + \frac{\zeta \Delta \mu}{(\Gamma \gamma_{22} - \gamma_{12}^2)} (\gamma_{22} \dot{X} x - \gamma_{12} \dot{x} x) \right] \\ &= \lim_{t \rightarrow \infty} \frac{1}{t} \frac{1}{T} \int dt \left[ - \dot{H} + \frac{\zeta \Delta \mu}{(\Gamma \gamma_{22} - \gamma_{12}^2)} (\gamma_{22} \dot{X} x - \gamma_{12} \dot{x} x) \right] = \frac{\gamma_{22} \zeta \Delta \mu}{T(\Gamma \gamma_{22} - \gamma_{12}^2)} \langle \dot{X} x \rangle \end{aligned} \tag{E.5}$$

where the last line replaces a time-average by an ensemble average.

For (ii), the time-reversed action is

$$\begin{aligned} \mathcal{A}_{(ii)}^R &= \frac{1}{4T(\Gamma \gamma_{22} - \gamma_{12}^2)} \int dt \gamma_{22} \left[ -\dot{X} + \Gamma \frac{\partial H}{\partial X} - \gamma_{12} \frac{\partial H}{\partial x} + \zeta \Delta \mu x \right]^2 + \Gamma \left[ \dot{x} - \gamma_{22} \frac{\partial H}{\partial x} + \gamma_{12} \frac{\partial H}{\partial X} \right]^2 \\ &\quad - 2\gamma_{12} \left[ -\dot{X} + \Gamma \frac{\partial H}{\partial X} - \gamma_{12} \frac{\partial H}{\partial x} + \zeta \Delta \mu x \right] \left[ \dot{x} - \gamma_{22} \frac{\partial H}{\partial x} + \gamma_{12} \frac{\partial H}{\partial X} \right]. \end{aligned} \tag{E.6}$$

The ‘entropy production rate’ can be formally calculated using the earlier definition. Note, this should be simply interpreted as the ratio of probabilities of the forward process and a backward process conditioned on  $x$  being reversed.

$$\sigma_{(ii)} = \lim_{t \rightarrow \infty} \frac{1}{t} \frac{1}{T} \int dt \left[ - \left( \dot{X} \frac{\partial H}{\partial X} + \dot{x} \frac{\partial H}{\partial x} \right) + \zeta \Delta \mu \frac{\partial H}{\partial X} x - \gamma_{12} \frac{\partial H}{\partial X} \frac{\partial H}{\partial x} + \frac{\gamma_{12}}{(\Gamma \gamma_{22} - \gamma_{12}^2)} (\dot{X} \dot{x} - \zeta \Delta \mu \dot{x} x) \right]. \tag{E.7}$$

Taking the limit and converting time-averages to ensemble averages we obtain

$$\sigma_{(ii)} = \frac{\zeta \Delta \mu}{T} \left\langle \frac{\partial H}{\partial X} x \right\rangle + \gamma_{12} \left( \frac{1}{(\Gamma \gamma_{22} - \gamma_{12}^2)} \langle \dot{X} \dot{x} \rangle - \left\langle \frac{\partial H}{\partial X} \frac{\partial H}{\partial x} \right\rangle \right). \tag{E.8}$$

Note, as expected  $\sigma_{(ii)}$ , unlike  $\sigma_{(i)}$ , does not vanish in the  $\Delta \mu \rightarrow 0$  limit. It is nonetheless a measure of departure from thermal equilibrium, but the *extra* parameter governing that departure, when  $\Delta \mu = 0$ , is the difference between the value  $\gamma_{12}$  assigned to the off-diagonal kinetic coefficient  $\gamma_{21}$ , and the value it should have, namely,  $-\gamma_{12}$ . In the special case when  $\gamma_{12} = 0$ , we obtain the more familiar forms  $\sigma_{(i)} = \zeta \Delta \mu \langle \dot{X} x \rangle / T\Gamma$  and  $\sigma_{(ii)} = \zeta \Delta \mu \langle (\partial H / \partial X) x \rangle / T$ , both of which vanish in the  $\Delta \mu \rightarrow 0$  limit. The sum of these two is, as expected,  $(\zeta \Delta \mu)^2 \langle x^2 \rangle / T\Gamma$ . Thus,  $\Delta \mu = 0$  leads to two different equally physically meaningful equilibrium limits only when  $\Delta \mu$  performs a double duty – breaking time-reversal symmetry and providing the only coupling between  $X$  and  $x$ .

When the passive couplings between  $X$  and  $x$  are anti-symmetric instead, we have the set of equations

$$\dot{X} = -\Gamma \frac{\partial H}{\partial X} - \gamma_{12} \frac{\partial H}{\partial x} + \zeta \Delta \mu x + \xi_X \tag{E.9}$$

$$\dot{x} = -\gamma_{22} \frac{\partial H}{\partial x} + \gamma_{12} \frac{\partial H}{\partial X} + \xi_x \tag{E.10}$$

with the noise correlators  $\langle \xi_X(t) \xi_X(t') \rangle = 2T\Gamma \delta(t - t')$ ,  $\langle \xi_x(t) \xi_x(t') \rangle = 2T\gamma_{22} \delta(t - t')$ ,  $\langle \xi_X(t) \xi_x(t') \rangle = 0$ . The stochastic Onsager–Machlup action is

$$\mathcal{A} = \frac{1}{4T} \int dt \frac{1}{\Gamma} \left[ \dot{X} + \Gamma \frac{\partial H}{\partial X} + \gamma_{12} \frac{\partial H}{\partial x} - \zeta \Delta \mu x \right]^2 + \frac{1}{\gamma_{22}} \left[ \dot{x} + \gamma_{22} \frac{\partial H}{\partial x} - \gamma_{12} \frac{\partial H}{\partial X} \right]^2. \tag{E.11}$$

The first time reversal protocol is (i)  $t \rightarrow -t$ ,  $X \rightarrow X$ ,  $x \rightarrow x$  which results in the time-reversed action

$$\mathcal{A}_{(i)}^R = \frac{1}{4T} \int dt \frac{1}{\Gamma} \left[ -\dot{X} + \Gamma \frac{\partial H}{\partial X} + \gamma_{12} \frac{\partial H}{\partial x} - \zeta \Delta \mu x \right]^2 + \frac{1}{\gamma_{22}} \left[ -\dot{x} + \gamma_{22} \frac{\partial H}{\partial x} - \gamma_{12} \frac{\partial H}{\partial X} \right]^2 \quad (\text{E.12})$$

leading to the entropy production

$$\sigma_{(i)} = \lim_{t \rightarrow \infty} \frac{1}{t} \frac{1}{T} \int \left[ - \left( \dot{X} \frac{\partial H}{\partial X} + \dot{x} \frac{\partial H}{\partial x} \right) - \gamma_{12} \left( \frac{1}{\Gamma} \dot{X} \frac{\partial H}{\partial x} - \frac{1}{\gamma_{22}} \dot{x} \frac{\partial H}{\partial X} \right) + \frac{\zeta \Delta \mu}{\Gamma} \dot{X} x \right]. \quad (\text{E.13})$$

Again, assuming ergodicity,

$$\sigma_{(i)} = -\frac{\gamma_{12}}{T} \left\langle \frac{1}{\Gamma} \dot{X} \frac{\partial H}{\partial x} - \frac{1}{\gamma_{22}} \dot{x} \frac{\partial H}{\partial X} \right\rangle + \frac{\zeta \Delta \mu}{T\Gamma} \langle \dot{X} x \rangle \quad (\text{E.14})$$

is shown to be non zero in the limit  $\Delta \mu \rightarrow 0$  as expected and goes to 0 if  $\gamma_{12} = 0$ . Similarly, for the time-reversal operation (ii)  $t \rightarrow -t$ ,  $X \rightarrow X$ ,  $x \rightarrow -x$ ,

$$\mathcal{A}_{(ii)}^R = \frac{1}{4T} \int dt \frac{1}{\Gamma} \left[ -\dot{X} + \Gamma \frac{\partial H}{\partial X} - \gamma_{12} \frac{\partial H}{\partial x} + \zeta \Delta \mu x \right]^2 + \frac{1}{\gamma_{22}} \left[ \dot{x} - \gamma_{22} \frac{\partial H}{\partial x} - \gamma_{12} \frac{\partial H}{\partial X} \right]^2 \quad (\text{E.15})$$

leading to the entropy production

$$\sigma_{(ii)} = \lim_{t \rightarrow \infty} \frac{1}{t} \frac{1}{T} \int \left[ - \left( \dot{X} \frac{\partial H}{\partial X} + \dot{x} \frac{\partial H}{\partial x} \right) + \zeta \Delta \mu \frac{\partial H}{\partial X} x \right] = \frac{\zeta \Delta \mu}{T} \left\langle \frac{\partial H}{\partial X} x \right\rangle. \quad (\text{E.16})$$

As expected, this vanishes when  $\Delta \mu \rightarrow 0$ . Thus, when the passive dynamics is not  $x \rightarrow -x$  invariant, only one ‘time-reversal prescription’ vanishes in the  $\Delta \mu \rightarrow 0$  limit and it follows that it is the only admissible prescription.

### Appendix F. Nonequilibrium fluctuation-dissipation relation for an active particle in a harmonic trap

A pure AOUP (i.e. one without translational diffusion) in a harmonic potential is an equilibrium problem with an effective temperature

$$T_{\text{eff}} = T \frac{v^2 \gamma}{k(\Gamma k + K \gamma)}. \quad (\text{F.1})$$

In the case of a harmonic potential, we can calculate the covariance matrix from (8) and (9) as

$$\mathbf{C} = T_{\text{eff}} \begin{pmatrix} \frac{(1+\beta)k(\Gamma k + K \gamma) + v^2 \gamma}{\gamma K v^2} & \frac{1}{v} \\ \frac{1}{v} & \frac{\Gamma(\Gamma k + K \gamma)}{\gamma v^2} \end{pmatrix}. \quad (\text{F.2})$$

The steady-state distribution function is  $\propto e^{(-\phi/T_{\text{eff}})}$  where the potential

$$\phi = \frac{1}{2} \mathbf{q} \cdot \mathbf{U} \cdot \mathbf{q} \quad (\text{F.3})$$

with  $\mathbf{U} = T_{\text{eff}} \mathbf{C}^{-1}$  and  $\mathbf{q} = (X, x)$ . We now rewrite the dynamical equations in the form

$$\dot{\mathbf{q}} = -\mathbf{L} \cdot \frac{\partial \phi}{\partial \mathbf{q}} + \boldsymbol{\eta} \tag{F.4}$$

where  $\langle \boldsymbol{\eta}(t) \boldsymbol{\eta}(t') \rangle = 2\mathbf{D}T_{\text{eff}}\delta(t - t')$  with

$$\mathbf{D} = \begin{pmatrix} \frac{k(1+\beta)(\Gamma k + K\gamma)}{v^2\Gamma\gamma} & 0 \\ 0 & \frac{k(\Gamma k + K\gamma)\Gamma}{\gamma^2 v^2} \end{pmatrix} \quad \text{and } \mathbf{L} = -\mathbf{F} \cdot \mathbf{C}/T_{\text{eff}} \quad \text{where } \mathbf{F} = \begin{pmatrix} -\frac{K}{\Gamma} & \frac{v}{\Gamma} \\ 0 & -\frac{k}{\gamma} \end{pmatrix}.$$

Note that  $\mathbf{L} + \mathbf{L}^T = 2\mathbf{D}$ .

To calculate the response function we now modify  $\phi$ :  $\phi^h = \phi^0 - \mathbf{h} \cdot \mathbf{x}$ . We define the response function as

$$\mathcal{R}_{ij} = \left. \frac{\delta x_i}{\delta h_j} \right|_{h_i=0} = (-i\omega \mathbf{I} - \mathbf{F})^{-1} \cdot \mathbf{L} \tag{F.5}$$

where  $\mathbf{I}$  is the identity. Then,

$$T_{\text{eff}}i[\mathcal{R}_{ji}(-\omega) - \mathcal{R}_{ij}(\omega)] = 2T_{\text{eff}}\chi''_{ij}(\omega) = \omega \mathcal{C}_{ij}(\omega) \tag{F.6}$$

where  $\mathcal{C}_{ij}$  is the correlation matrix (note that  $\int (d\omega/2\pi)\mathcal{C}(\omega) = \mathbf{C}$ ). This relation holds for arbitrary  $\beta$  and, in particular, also for  $\beta = -1$  which corresponds to the zero translational diffusion case. For  $\beta = -1$ , the equation of motion becomes

$$\frac{\Gamma\gamma}{k}\ddot{X} = -\left(\frac{K\gamma}{k} + \Gamma\right)\dot{X} - KX + \frac{v}{k}\xi^p. \tag{F.7}$$

Adding a perturbing force to (F.7) results in the response function with an imaginary part

$$\tilde{\chi}''_{XX0}(\omega) = \frac{k(k\Gamma + K\gamma)\omega}{(K^2 + \Gamma^2\omega^2)(k^2 + \gamma^2\omega^2)} \tag{F.8}$$

and the fluctuation-dissipation theorem  $2T_{\text{eff}}\tilde{\chi}''_{XX0}(\omega) = \omega C_{XX0}(\omega)$ .

Note that despite the formal similarity to an equilibrium FDT, there is an important difference: in equilibrium, we would have added the perturbation directly to the Hamiltonian i.e. the perturbing force would have appeared directly in original equations of motion. We do not do that here. The force that appears in the equation of motion is  $\mathbf{L} \cdot \mathbf{h}$ . However, the response function is defined not with respect to this force but purely with respect to  $\mathbf{h}$ . This FDT was first explicitly calculated by Eyink, Lebowitz and Spohn [84], inspired by a more general result by Graham [85]. Basically, this finds a force the response to which equals the correlation function in the steady state. This is distinct from the fluctuation-dissipation relation discussed in [86] which amounts to adding perturbing forces to *original* equations of motion and finding the correlation functions that the response to these forces are equal to. For linear dynamics, these variables whose correlators correspond to the response are given for our system as  $(\mathbf{F}^{-1})^T \cdot \mathbf{U} \cdot \mathbf{x}$ .

Finally, one might argue that the strength of noise should be immaterial for the response to an external physical force. Despite the response functions we calculate not being responses to physical external forces, let me examine this. Basically, let us calculate the response to a perturbation when the translational diffusion is 0 i.e. use

(F.8) as the response function even in the presence of translational diffusion. The  $XX$  correlation function in the presence of translational diffusion has the value

$$\mathcal{C}_{XX}(\omega) = 2T \frac{v^2\gamma + (1 + \beta)[\Gamma(k^2 + \gamma^2\omega^2)]}{(K^2 + \Gamma^2\omega^2)(k^2 + \gamma^2\omega^2)}. \tag{F.9}$$

This implies

$$\mathcal{C}_{XX}(\omega) - \frac{2T_{\text{eff}}}{\omega} \tilde{\chi}''_{XX0}(\omega) = \frac{1 + \beta}{\Gamma} \frac{2T}{\omega^2 + K^2/\Gamma^2}. \tag{F.10}$$

This implies that at short times i.e. when  $\omega \gg K/\Gamma$ , the measured correlation function will seem to satisfy a fluctuation-dissipation relation with the response function measured at 0 translational diffusion, even when translational diffusion is present. This denotes the time-scale below which AOUP seems to have inertia even in the presence of translational diffusion.

**Appendix G. Harada–Sasa relation for active field theory of polar rods**

In this appendix, we derive the Harada–Sasa relation for the field theory of active polar rods. To calculate the response functions, we add  $-h_1\rho$  to  $F_\rho$  and  $-\mathbf{h}_2 \cdot \mathbf{p}$  to  $F_{\mathbf{p}}$ . The action becomes

$$\mathcal{A}^{OM} = \frac{1}{4} \int \left[ -\frac{1}{T} \left\{ \dot{\rho} + v\nabla \cdot (\rho\mathbf{p}) - \nabla^2 \frac{\delta F}{\delta \rho} + \nabla^2 h_1 \right\} \nabla^{-2} \left\{ \dot{\rho} + v\nabla \cdot (\rho\mathbf{p}) - \nabla^2 \frac{\delta F}{\delta \rho} + \nabla^2 h_1 \right\} + \frac{1}{T} \left( \dot{\mathbf{p}} + \lambda\mathbf{p} \cdot \nabla\mathbf{p} + \frac{\delta F}{\delta \mathbf{p}} - \mathbf{h}_2 \right)^2 \right] \tag{G.1}$$

with the part linear in  $h_1$  and  $\mathbf{h}_2$  being

$$\delta\mathcal{A}^{OM} = -\frac{1}{2} \int \left[ \frac{1}{T} h_1 \left\{ \dot{\rho} + v\nabla \cdot (\rho\mathbf{p}) - \nabla^2 \frac{\delta F}{\delta \rho} \right\} + \frac{1}{T} \mathbf{h}_2 \cdot \left( \dot{\mathbf{p}} + \lambda\mathbf{p} \cdot \nabla\mathbf{p} + \frac{\delta F}{\delta \mathbf{p}} \right) \right]. \tag{G.2}$$

Directly evaluating the response functions from  $\delta\mathcal{A}^{OM}$  we find

$$\begin{aligned} \mathcal{R}_{\rho\rho}(t) - \mathcal{R}_{\rho\rho}(-t) = & -\frac{1}{T} \partial_t \langle \rho(\mathbf{x}, t)\rho(\mathbf{x}, 0) \rangle + \frac{1}{2T} \left\langle \rho(\mathbf{x}, t) \left[ v\nabla \cdot (\rho\mathbf{p}) - \nabla^2 \frac{\delta F}{\delta \rho} \right] (\mathbf{x}, 0) \right. \\ & \left. - \left[ v\nabla \cdot (\rho\mathbf{p}) - \nabla^2 \frac{\delta F}{\delta \rho} \right] (\mathbf{x}, t) \rho(\mathbf{x}, 0) \right\rangle \end{aligned} \tag{G.3}$$

which implies

$$\begin{aligned} T[\mathcal{R}_{\rho\rho}(t) - \mathcal{R}_{\rho\rho}(-t)] + \partial_t C_{\rho\rho} = & \frac{1}{2} \left\langle \rho(\mathbf{x}, t) \left[ v\nabla \cdot (\rho\mathbf{p}) - \nabla^2 \frac{\delta F}{\delta \rho} \right] (\mathbf{x}, 0) \right. \\ & \left. - \left[ v\nabla \cdot (\rho\mathbf{p}) - \nabla^2 \frac{\delta F}{\delta \rho} \right] (\mathbf{x}, t) \rho(\mathbf{x}, 0) \right\rangle. \end{aligned} \tag{G.4}$$

Similarly,

$$T[\mathcal{R}_{pp}(t) - \mathcal{R}_{pp}(-t)] + \partial_t C_{pp} = \frac{1}{2} \left\langle \mathbf{p}(\mathbf{x}, t) \cdot \left( \lambda \mathbf{p} \cdot \nabla \mathbf{p} + \frac{\delta F}{\delta \mathbf{p}} \right) (\mathbf{x}, 0) - \left( \lambda \mathbf{p} \cdot \nabla \mathbf{p} + \frac{\delta F}{\delta \mathbf{p}} \right) (\mathbf{x}, t) \cdot \mathbf{p}(\mathbf{x}, 0) \right\rangle. \quad (\text{G.5})$$

The entropy production rate can now be expressed in terms of the response and correlation functions using the definition and the expression for the entropy production rate in section 4.2.1:

$$\sigma = \frac{1}{T} \lim_{t \rightarrow 0} \int d\mathbf{x} \partial_t \left[ \nabla^{-2} (T[\mathcal{R}_{\rho\rho}(t) - \mathcal{R}_{\rho\rho}(-t)] + \partial_t C_{\rho\rho}) - (T[\mathcal{R}_{pp}(t) - \mathcal{R}_{pp}(-t)] + \partial_t C_{pp}) \right]. \quad (\text{G.6})$$

This is the Harada–Sasa relation for the Toner–Tu model. Defining  $[\mathcal{R}_{\rho\rho}(t) - \mathcal{R}_{\rho\rho}(-t)] = 2i\chi''_{\rho\rho}(\mathbf{q}, \omega)$  and  $[\mathcal{R}_{pp}(t) - \mathcal{R}_{pp}(-t)] = 2i\chi''_{pp}(\mathbf{q}, \omega)$ , where  $\mathbf{q}$  is the wavevector, we find an expression for the frequency and wavevector resolved entropy production  $\sigma = \int d\omega d\mathbf{q} \sigma_{\omega\mathbf{q}}$ ,

$$\sigma_{\omega\mathbf{q}} = \frac{\omega}{(2\pi)^{d+1} T} \left[ q^2 \{ \omega C_{\rho\rho}(\mathbf{q}, \omega) - 2T\chi''_{\rho\rho}(\mathbf{q}, \omega) \} + \{ \omega C_{pp}(\mathbf{q}, \omega) - 2T\chi''_{pp}(\mathbf{q}, \omega) \} \right]. \quad (\text{G.7})$$

This expression opens up the possibility of quantitatively measuring the departure of a flock from equilibrium through difficult, but in principle feasible correlation and response measurements.

### Appendix H. Entropy production in the homogeneous polarised phase in the $T \rightarrow 0$ limit

In active model B, entropy production is small within homogeneous phases. We now show that this is not the case for the Toner–Tu model. We expand both  $\rho$  and  $\mathbf{p}$  in noise-strength in the weak-noise limit.  $\rho = \rho_0 + \sqrt{T}\rho_1 + T\rho_2 + \dots$ ; and  $\mathbf{p} = \mathbf{p}_0 + \sqrt{T}\mathbf{p}_1 + T\mathbf{p}_2 + \dots$ . Assuming (though the form of the free-energy does not matter)

$$F = \int d\mathbf{x} \left[ \frac{A}{2} \rho^2 + \frac{\alpha}{2} p^2 + \frac{\beta}{4} p^4 + \ell \mathbf{p} \cdot \nabla \rho + \frac{K}{2} (\nabla \mathbf{p})^2 \right] \quad (\text{H.1})$$

we obtain the equations of motion

$$\dot{\rho}_0 = -v \nabla \cdot (\rho_0 \mathbf{p}_0) + \nabla^2 \mu_0 \quad (\text{H.2})$$

$$\dot{\rho}_1 = -v \nabla \cdot (\rho_1 \mathbf{p}_0 + \rho_0 \mathbf{p}_1) + \nabla^2 \mu_1 + \xi \quad (\text{H.3})$$

$$\dot{\mathbf{p}}_0 + \lambda \mathbf{p}_0 \cdot \nabla \mathbf{p}_0 = -\mathbf{H}_0 \quad (\text{H.4})$$

$$\dot{\mathbf{p}}_1 + \lambda (\mathbf{p}_1 \cdot \nabla \mathbf{p}_0 + \mathbf{p}_0 \cdot \nabla \mathbf{p}_1) = -\mathbf{H}_1 + \boldsymbol{\xi}^p \quad (\text{H.5})$$

where  $\mathbf{H}_0 = \alpha \mathbf{p}_0 + \beta p^2 \mathbf{p}_0 + \ell \nabla \rho_0 - K \nabla^2 \mathbf{p}_0$ ,  $\mathbf{H}_1 = \alpha \mathbf{p}_1 + 3\beta p^2 \mathbf{p}_1 + \ell \nabla \rho_1 - K \nabla^2 \mathbf{p}_1$ ,  $\mu_0 = A\rho_0 - \ell \nabla \cdot \mathbf{p}_0$ , and  $\mu_1 = A\rho_1 - \ell \nabla \cdot \mathbf{p}_1$ . The action, to zeroth order in noise-strength, is



$$\mathcal{A}^{OM} = \int dt d\mathbf{x} [-\{\dot{\rho}_1 + v \nabla \cdot (\rho_1 \mathbf{p}_0 + \rho_0 \mathbf{p}_1) - \nabla^2 \mu_1\} \nabla^{-2} \{\dot{\rho}_1 + v \nabla \cdot (\rho_1 \mathbf{p}_0 + \rho_0 \mathbf{p}_1) - \nabla^2 \mu_1\} + \{\dot{\mathbf{p}}_1 + \lambda(\mathbf{p}_1 \cdot \nabla \mathbf{p}_0 + \mathbf{p}_0 \cdot \nabla \mathbf{p}_1) + \mathbf{H}_1\}^2]. \quad (\text{H.6})$$

We now assume that both  $\rho_0$  and  $\mathbf{p}_0$  relaxes to constant profiles. Even then, there is a part of entropy production rate that is independent of noise strength:

$$\sigma_0 = v \langle \dot{\rho}_1 \nabla^{-2} (\rho_0 \nabla \cdot \mathbf{p}_1 + \mathbf{p}_0 \cdot \nabla \rho_1) - \lambda \langle \dot{\mathbf{p}}_1 \cdot (\mathbf{p}_0 \cdot \nabla) \mathbf{p}_1 \rangle + O(\sqrt{T}). \quad (\text{H.7})$$

This implies that even homogeneous phases have an entropy-production rate at  $O(T_1^0)$  unlike in active scalar models (model *A* or *B*). This is to be expected on symmetry grounds.

## References

- [1] Ramaswamy S 2010 *Ann. Rev. Condens. Matter Phys.* **1** 323
- [2] Marchetti M C, Joanny J F, Ramaswamy S, Liverpool T B, Prost J, Rao M and Simha R A 2013 *Rev. Mod. Phys.* **85** 1143
- [3] Fodor É, Nardini C, Cates M E, Tailleur J, Visco P and van Wijland F 2016 *Phys. Rev. Lett.* **117** 038103
- [4] Jülicher F, Ajdari A and Prost J 1997 *Rev. Mod. Phys.* **69** 1269
- [5] Kruse K, Joanny J F, Jülicher F, Prost J and Sekimoto K 2005 *Eur. Phys. J. E* **16** 5
- [6] Joanny J F, Jülicher F, Kruse K and Prost J 2007 *New J. Phys.* **9** 422
- [7] Jülicher F, Kruse K, Prost J and Joanny J F 2007 *Phys. Rep.* **449** 3
- [8] Jülicher F, Grill S W and Salbreux G 2018 *Rep. Prog. Phys.* **81**
- [9] Pietzonka P and Seifert U 2017 *J. Phys. A: Math. Theor.* **51** 01LT01
- [10] Neri I, Roldán É and Jülicher F 2017 *Phys. Rev. X* **7** 011019
- [11] Shankar S and Marchetti M C 2018 *Phys. Rev. E* **98** 020604
- [12] Pigolotti S, Neri I, Roldán É and Jülicher F 2017 *Phys. Rev. Lett.* **119** 140604
- [13] Kumar N, Soni H, Ramaswamy S and Sood A K 2015 *Phys. Rev. E* **91** 030102
- [14] Schimansky-Geier L, Mieth M, Rosé H and Malchow H 1995 *Phys. Lett. A* **207** 140
- [15] Ebeling W, Schweitzer F and Tilch B 1999 *BioSystems* **49** 17
- [16] Romanczuk P, Bär M, Ebeling W, Lindner B and Schimansky-Geier L 2012 *Eur. Phys. J. Spec. Top.* **202** 1
- [17] Maggi C, Paoluzzi M, Pellicciotta N, Lepore A, Angelani L and Di Leonardo R 2014 *Phys. Rev. Lett.* **113** 238303
- [18] Koumakis N, Maggi C and Di Leonardo R 2014 *Soft Matter* **10** 5695
- [19] Toner J and Tu Y 1995 *Phys. Rev. Lett.* **75** 4326
- [20] Toner J and Tu Y 1998 *Phys. Rev. E* **58** 4828
- [21] Toner J, Tu Y and Ramaswamy S 2005 *Ann. Phys.* **318** 170
- [22] Sandford C, Grosberg A Y and Joanny J F 2017 *Phys. Rev. E* **96** 052605
- [23] Harada T and Sasa S I 2005 *Phys. Rev. Lett.* **95** 130602
- [24] Wang S W, Kawaguchi K, Sasa S I and Tang L H 2016 *Phys. Rev. Lett.* **117** 070601
- [25] Brand H R, Pleiner H and Sven vsek D 2014 *Eur. Phys. J. E* **37** 83
- [26] Ramaswamy S 2017 *J. Stat. Mech.* **054002**
- [27] Kumar K V, Ramaswamy S and Rao M 2008 *Phys. Rev. E* **77** 020102
- [28] Ma S K and Mazenko G F 1975 *Phys. Rev. B* **11** 4077
- [29] Chaikin P M and Lubensky T C 1995 *Principles of Condensed Matter Physics* vol 1 (Cambridge: Cambridge University Press)
- [30] Zwanzig R 2001 *Nonequilibrium Statistical Mechanics* (Oxford: Oxford University Press)
- [31] Lau A W and Lubensky T C 2007 *Phys. Rev. E* **76** 011123
- [32] Casimir H B G 1945 *Rev. Mod. Phys.* **17** 343
- [33] De Groot S R and Mazur P 2013 *Non-Equilibrium Thermodynamics* (North Chelmsford, MA: Courier Corporation)
- [34] Mori H 1965 *Prog. Theor. Phys.* **33** 423
- [35] Mazenko G F 2008 *Nonequilibrium Statistical Mechanics* (New York: Wiley)

- [36] Hohenberg P C and Halperin B I 1977 *Rev. Mod. Phys.* **49** 435
- [37] Baule A, Kumar K V and Ramaswamy S 2008 *J. Stat. Mech.* P11008
- [38] Uhlenbeck G E and Ornstein L S 1930 *Phys. Rev.* **36** 823
- [39] Ramaswamy S, Toner J and Prost J 2000 *Phys. Rev. Lett.* **84** 3494
- [40] Prost J and Bruinsma R 1996 *Europhys. Lett.* **33** 321
- [41] Jung P and Hänggi P 1987 *Phys. Rev. A* **35** 4464
- [42] Hänggi P and Jung P 1995 *Adv. Chem. Phys.* **89** 239
- [43] Toner J 2012 *Phys. Rev. E* **86** 031918
- [44] Landau L and Lifshitz E 1959 *Course of Theoretical Physics. Vol. 6: Fluid Mechanics* (Elsevier: London)
- [45] Brotto T, Caussin J B, Lauga E and Bartolo D 2013 *Phys. Rev. Lett.* **110** 038101
- [46] Kumar N, Soni H, Ramaswamy S and Sood A 2014 *Nat. Commun.* **5** 4688
- [47] Kung W, Marchetti M C and Saunders K 2006 *Phys. Rev. E* **73** 031708
- [48] Erdmann U, Ebeling W, Schimansky-Geier L and Schweitzer F 2000 *Eur. Phys. J. B* **15** 105
- [49] Schweitzer F, Ebeling W and Tilch B 1998 *Phys. Rev. Lett.* **80** 5044
- [50] Simha R A and Ramaswamy S 2002 *Phys. Rev. Lett.* **89** 058101
- [51] Gruler H, Dewald U and Eberhardt M 1999 *Eur. Phys. J. B* **11** 187
- [52] Mishra S 2010 *PhD Thesis* Indian Institute of Science
- [53] Ramaswamy S, Simha R A and Toner J 2003 *Europhys. Lett.* **62** 196
- [54] Mani R G, Smet J H, von Klitzing K, Narayanamurti V, Johnson W B and Umansky V 2002 *Nature* **420** 646
- [55] Zudov M, Du R, Pfeiffer L and West K 2003 *Phys. Rev. Lett.* **90** 046807
- [56] Alicea J, Balents L, Fisher M P, Paramakanti A and Radzihovsky L 2005 *Phys. Rev. B* **71** 235322
- [57] Dunkel J, Heidenreich S, Drescher K, Wensink H H, Bär M and Goldstein R E 2013 *Phys. Rev. Lett.* **110** 228102
- [58] Swift J and Hohenberg P C 1977 *Phys. Rev. A* **15** 319
- [59] Ramaswamy S and Mazenko G F 1982 *Phys. Rev. A* **26** 1735
- [60] Kim B and Mazenko G F 1991 *J. Stat. Phys.* **64** 631
- [61] Toner J 2012 *Phys. Rev. Lett.* **108** 088102
- [62] Lebowitz J L and Spohn H 1999 *J. Stat. Phys.* **95** 333
- [63] Kullback S and Leibler R A 1951 *Ann. Math. Stat.* **22** 79
- [64] Nardini C, Fodor É, Tjhung E, Van Wijland F, Tailleur J and Cates M E 2017 *Phys. Rev. X* **7** 021007
- [65] Onsager L and Machlup S 1953 *Phys. Rev.* **91** 1505
- [66] Cugliandolo L F and Lecomte V 2017 *J. Phys. A: Math. Theor.* **50** 345001
- [67] Teramoto H and Sasa S I 2005 *Phys. Rev. E* **72** 060102
- [68] Harada T and Sasa S I 2006 *Phys. Rev. E* **73** 026131
- [69] Yamada K and Yoshimori A 2015 *J. Phys. Soc. Japan* **84** 044008
- [70] Martin P C, Siggia E D and Rose H A 1973 *Phys. Rev. A* **8** 423
- [71] Janssen H K 1976 *Z. Phys. B* **23** 377
- [72] Janssen H 1979 *Dynamical Critical Phenomena and Related Topics* (Berlin: Springer) pp 25–47
- [73] De Dominicis C 1976 *J. Phys. Colloques* **37** C1–247
- [74] De Dominicis C 1978 *Phys. Rev. B* **18** 4913
- [75] Aron C, Biroli G and Cugliandolo L F 2010 *J. Stat. Mech.* P11018
- [76] Grosberg A and Joanny J F 2015 *Phys. Rev. E* **92** 032118
- [77] Saha S, Golestanian R and Ramaswamy S 2014 *Phys. Rev. E* **89** 062316
- [78] Liebchen B, Marenduzzo D and Cates M E 2017 *Phys. Rev. Lett.* **118** 268001
- [79] Liebchen B, Cates M E and Marenduzzo D 2016 *Soft Matter* **12** 7259
- [80] Gladrow J, Fakhri N, MacKintosh F, Schmidt C and Broedersz C 2016 *Phys. Rev. Lett.* **116** 248301
- [81] Battle C, Broedersz C P, Fakhri N, Geyer V F, Howard J, Schmidt C F and MacKintosh F C 2016 *Science* **352** 604
- [82] Seifert U 2012 *Rep. Prog. Phys.* **75** 126001
- [83] Rytter D 1981 *Z. Phys. B* **41** 39
- [84] Eyink G L, Lebowitz J L and Spohn H 1996 *J. Stat. Phys.* **83** 385
- [85] Graham R 1977 *Z. Phys. B* **26** 397
- [86] Prost J, Joanny J F and Parrondo J 2009 *Phys. Rev. Lett.* **103** 090601